



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	重庆交通大学
参赛队号	19106180041
队员姓名	1.侯雨彤
	2.王锡峰
	3.青 宇

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 全球变暖和极端气候的时空分析

摘 要：

本文以 **LSTM 神经网络** 和学习算法为基础，针对全球变暖现象进行了量化分析和研究，在对各类数据样本进行预处理的时候，运用了均一性检验和质量控制以及相邻线性插值的方法。运用**多元高次多项式的非线性回归**以及“**小波分析—越险指标**”和“**上述多因素决策气候指标**”等方法建立分析数学模型，利用 python 和 matlab 编程软件对上述数学模型进行程序实现了对于全球变暖现象的量化分析和研究。

针对问题一：对收集的数据进行严格的质量控制和均一性分析处理并且针对部分缺失值进行了线性插值处理，建立了 LSTM 时空序列数据分析模型在空间多测度、时间多时相分析天气变化历史数据的时空趋势及规律，得出了随着全球气候变暖出现极端气候的结论以及通过分析发现如下结果：（1）加拿大全国年平均气温在统计的 26 年间温度整体上升 1.2℃；（2）加拿大全国低纬度地区温度上升速率大于高纬度地区；（3）全球海洋温度趋势整体上升并且仍然会以 0.2℃/10 年的趋势上升。（4）基于海洋温度的变化可以得到出现厄尔尼诺现象的时空特征。

针对问题二：全面考虑地球吸热、散热，海洋温度变化数据以及地气系统平衡方程的基础上通过各指标的统计分析确定复杂气候模型系数，基于多元高次多项式非线性回归的方法对装配所得复杂气候模型进行简化并且得到各分项系数，比较简化前后的预测模型验证模型的准确性并且得出未来 25 年的预测规律得出了**全球气候是持续变暖**的结论以及得出如下结果：（1）预测在 2045 年左右全球平均气温将升高 0.6 摄氏度左右；（2）由于变暖预测显示出一种“厚尾”分布，2045 年后其继续变暖变暖的可能性更大。

针对问题三：梳理收集极端气象的统计数据，建立可表征极端气象在全球范围内发生规律的“小波分析—越险指标”天气预测修正模型，基于上述修正模型得出**持续变暖致使了极寒天气在频次、程度、概率的增加**的结论以及如下结果：（1）随着全球气温的上升，极寒天气的出现频率和剧烈程度均有所提高；（2）全球变暖与局部地区极寒现象并不矛盾。

针对问题四：以问题一中加拿大数据集和问题二三的数学模型作为参考分析了极端天气的频次、程度和概率指标，解释了全球变暖与今年冬天特别冷之间的关系，以“**多因素决策气候指标**”作为考虑了气候变化趋势和复杂性的“全球变暖”的新概念。

关键词：温室效应 极寒天气 多目标规划 LSTM 神经网络机器 小波分析—越险指标 多因素决策气候指标

目录

一、 问题重述	1
1.1 问题背景	1
1.2 问题提出	1
二、 问题分析	2
三、 基本假设及相关定义	4
3.1 基本假设	4
3.2 符号说明	4
四、 模型的建立与求解	5
4.1 问题一：基于 LSTM 的气候变化时空数据关键信息提取及趋势分析模型	5
4.1.1 数据来源及预处理	5
4.1.1.1 数据来源	5
4.1.1.2 数据的预处理的方式	5
4.1.2 时间空间输入序列数据集	8
4.1.3 基于 M-K 的非参数统计检验分析	9
4.1.4 LSTM 气候时空趋势变化模型的建立	11
4.1.5 模型的求解与结果	12
4.2 问题二：基于多元高次多项式非线性回归的简化气候预测模型	23
4.2.1 相关因素影响机理分析及权重估计	23
4.2.2 复杂气候模型—子模型的建立	26
4.2.2.1 海洋温度模型	26
4.2.2.2 二氧化碳含量模型	27
4.2.2.3 人口及经济发展模型	28
4.2.2.4 厄尔尼诺以及拉尼娜周期变化模型	28
4.2.3 复杂气候模型—子模型数据的收集与预处理	30
4.2.3.1 人口因素	30
4.2.3.2 GDP	31
4.2.3.3 CO ₂	33
4.2.3.4 森林覆盖率	34
4.2.4 基于能量守恒的多因素复杂气候模型的建立	34
4.2.5 基于多元高次多项式非线性回归的模型简化与求解	35
4.2.6 模型的求解与结果	36
4.3 问题三：考虑极端气候现象的“小波分析-越险指标”天气预测模型	38
4.3.1 极端天气机理分析	38
4.3.2 小波分析的极端气候现象周期趋势分析	39
4.3.2.1 确定小波变换参数	39
4.3.2.2 极寒天气周期预测函数	40
4.3.2.3 开关函数	40

4.3.3 基于越险风险指标的极寒天气阈值计算模型	41
4.3.4 “小波分析-越险指标”的极端气候现象天气预测修正模型	41
4.3.5 模型的求解与结果	41
4.4 问题四：表征全球变暖的新概念“多因素气候决策指标”的提出	42
4.4.1 分析数据	42
4.4.2 统计分析及结论	44
五 模型的评价	47
5.1 模型的优点	47
5.2 模型的缺点	47
六 参考文献	48
七 附录	50

一、 问题重述

1.1 问题背景

温室效应又称为“花房效应”，是大气保温效应的俗称，自工业革命以来，人类向大气中排放的二氧化碳等吸热性温室气体逐年增加，大气的温室效应也随之增强从而引起了全球气候变暖等一系列极其严重的问题，这些温室气体对来自太阳辐射的可见光具有高度透过性，而对地球发射出来的长波辐射具有高度吸收性，能强烈吸收地面辐射中的红外线，使得地球温度上升，即温室效应。由于存在温室效应，影响地气系统吸收与发射的能量平衡，能量不断在地气系统累积，从而导致温度上升，造成全球气候变暖。许多科学家认为，全球变暖可能导致更多的极端气象的产生^[4]，导致全球降水量重新分配、冰川和冻土消融、海平面上升等威胁人类生存的因素。不过，虽然温室气体的浓度在不断上升，但自从进入 21 世纪以来，10 年间全球全年平均气温上升率仅为 0.03℃，几乎未变化，这种现象叫作全球变暖停滞状态。正因为出现全球变暖停滞现象，使公众对全球变暖产生了怀疑。公众对于全球变暖产生怀疑的原因主要在于观察问题的角度和范围。今年的夏天特别热或今年的冬天特别冷是地球上局地人们的直接感受，是一种天气现象。天气是一定区域短时段内的大气状态（如冷暖、风雨、干湿、阴晴等）及其变化的总称。而气候则是长时间内气象要素和天气现象的平均或统计状态，时间尺度为月、季、年、数年到数十年。气候是长时间的平均状态，在短时间内变化不大，所以人们一般感受不到。全球变暖是在气候尺度上看全球问题。从气候角度研究全球温度变化需要全球范围长时间的观测积累，但过去这方面的时空数据并不完整，给统计计算带来极大困难。不仅如此，海洋吸收热量对全球气候变化的影响很大。观测发现海洋表面温度的变化具有某种震荡特征，如年代际太平洋震荡、厄尔尼诺现象、拉尼娜现象等。这些因素使得研究全球温度变化更加困难。

1.2 问题提出

本文以此为背景，主要考虑一下几个问题：（1）对于加拿大和全球海洋时空序列进行数据分析得出模型在空间多测度、时间多时相分析天气变化历史数据的时空趋势及规律；（2）建立考虑多因素的刻画未来 25 年的气候变化的预测数学模型；（3）建立考虑极寒天气的天气预测数学模型说明全球变暖与巨地极寒现象之间是否矛盾；（4）基于前文数据分析和预测模型反映的气候变化机理，说明全球变暖和人体实际感官之间的关系并且提出一种全新的概念来表明这种走势。

二、 问题分析

全球气候变暖影响世界范围内气候系统的动态平衡，可能出现越来越多极端气象，导致全球降水量的重新分配、冰川和冻土消融致使海平面上升等严重威胁人类生存。全球气候变暖的解释是由于温室效应不断积累所致^[1]，这些温室气体对来自太阳辐射的可见光具有高度透过性，而对地球发射出来的长波辐射具有高度吸收性，能强烈吸收地面辐射中的红外线，使得地球温度上升，即温室效应。综合的专业气候预测模型框架主要有最优化模型和 CGE 模型以及模拟模型三大类，上述模型的目标函数分为福利最大化和成本最小化，气候模型的求解以及分析的机器学习算法主要有：BP 神经网络；RNN 神经网络；LSTM 神经网络等。上述模型及算法在不同程度上可量化分析全球气候时空数据及其趋势，预测未来全球气候变化，反映全球气候变化态势，但仍存在一定的局限性：①全球变暖是在气候尺度上看全球问题，从气候角度研究全球温度变化需要全球范围长时间的观测积累，但过去这方面的时空数据并不完整，给统计计算带来极大困难。②海洋吸收热量对全球气候变化的影响很大^[2]。观测发现海洋表面温度的变化具有某种震荡特征，如年代际太平洋震荡、厄尔尼诺现象、拉尼娜现象等，这些因素使得研究全球温度变化更加困难。③复杂的专业气候模型对于非专业人士难以理解和认识全球气候变化的态势，难以意识到气候变化、极端气候与全球气候变暖的关联性。

基于以上研究，通过对前述问题，尤其是已有气候模型及全球气候态势的相关研究结论、条件与数据的初步分析，通过加拿大政府官方网站下载的加拿大全国温度数据以及世界银行公开数据查找相关经济数据并且通过线性差值等方法进行数据预处理，最终获取加拿大以及全球气候变化以及影响因子时空序列数据针对全球变暖四个问题分别进行了如下问题的转化、简化与分析：

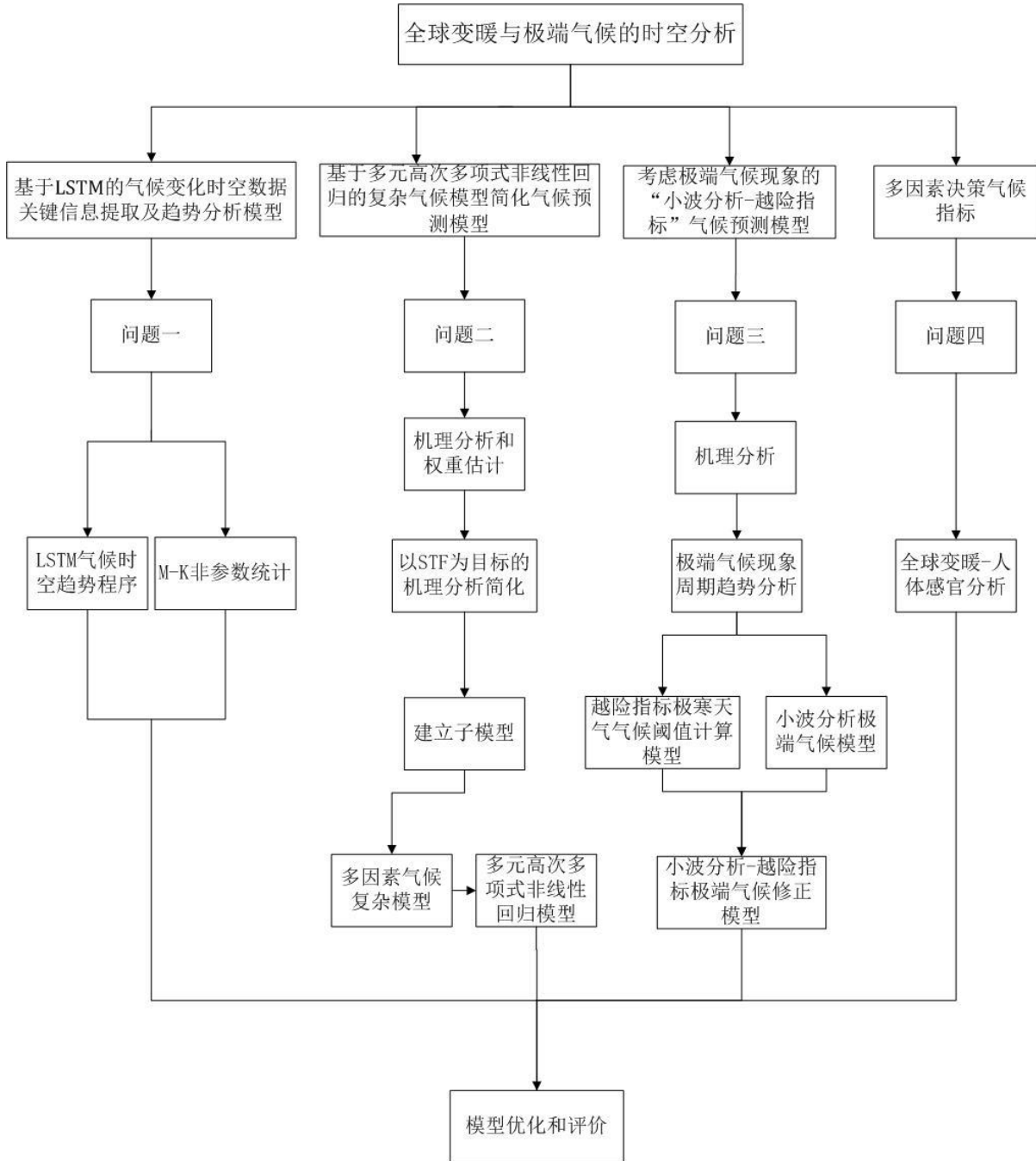
问题一：为提取、挖掘加拿大 178 个气温观测站近 25 年天气变化数据、海洋表面温度历史数据的时空变化趋势，对收集的数据进行严格的质量控制和均一性分析处理并且针对部分缺失值进行了线性插值处理，建立了 LSTM 时空序列数据分析模型在空间多测度、时间多时相分析天气变化历史数据的时空趋势及规律。

问题二：为建立刻画气候变化的模型并预测 25 年间的气候变化，但全球气候变化需要全球范围长时间的观测积累，气候变化涉及众多因素且数据并不完整，因此，在尽可能收集梳理全球范围内近 30 年的全球气候变化数据及影响因素的基础上，分析全球气候变化相关因素的影响机理，以 STF 目标函数进行了影响机理的多因素路径分析，略去权重高阶微量建立简化机理分析图，以简化机理分析图为基础构建影响因素子模型并且对子模型进行装配得到复杂气候模型，全面考虑地球吸热、散热，海洋温度变化数据以及地气系统平衡方程的基础上通过各指标的统计分析确定复杂气候模型系数，基于多元高次多项式非线性回归的方法对装配所得复杂气候模型进行简化并且得到各分项系数，比较简化前后的预测模型验证模型的准确性并且得出未来 25 年的预测规律。

问题三：为研究极端气象与气候变化的关系以及全球变暖与局部极寒现象的关系，在问题二数据及地气系统平衡模型的基础上，梳理收集极端气象的统计数据，建立可表征极端气象在全球范围内发生规律的“小波分析-越险指标”天气预测修正模型，通过该修正模型对问题二中的模型进行修正进而构建考虑极寒天气的全球气候互反馈综合模型反映整体气候-局部天气间的关系，通过对模型的量化求解综合分析极端气象的出现是否由气候变化引起，全球变暖与局部极寒现象的矛盾关系。

问题四：为进一步直观清晰反映全球变暖与局部极寒的关系，在问题二、三的简化气候模型与极端气候模型基础上，在时间维度与空间维度充分考虑全球变暖在全球范围大尺度与长时间的数据积累，针对气候变化在时空域的特性重新定义“全球变暖”，分析气候变化的趋势与复杂性，解释“全球变暖与某地今年冬天特别冷”的关系。

上述四个问题的解决全过程流程图如下所示：



三、基本假设及相关定义

3.1 基本假设

- A 假设在预测年限内，不会发生影响地球生态系统以及地月系统稳定的天体撞击事件。
- B 假设太阳活动稳定，太阳对地球提供的总热量恒定不变。
- C 假设宇宙沙尘浓度不发生变化，即太阳与地球之间热量传递介质理化性质不发生改变。
- D 假设地球磁场和轨道偏向率不发生变化。
- E 假设地球板块不发生移动。
- F 假设本文收集的数据资料真实可靠，符合数据归类分析的基本要求，能准确反应全球气候变化的基本规律。
- G 假设人类科技不会出现巨大突破，仍然以现有的能源为主
- H 假设未来全球范围内不会发生大规模的战争和社会动荡，造成人口大量变化。
- I 假设在各大网站查询到的数据都为真实数据。
- J 假设在未来 25 年内不会产生新的影响气候变化的因素。

3.2 符号说明

符号	符号含义
r_{kl}	待检站与临近站序列的相关系数
β	气温变化趋势大小指标
X_i	气温时间序列
Z	标准化统计量
\tilde{c}_t	t 时刻记忆细胞的更新状态
c_t	记忆细胞在 t 时刻的输出
i_t	输入门在 t 时刻的输出
f_t	遗忘门在 t 时刻的输出
o_t	输出门在 t 时刻的输出
h_t	隐藏层在 t 时刻的输出
σ	sigmoid 激活函数

四、模型的建立与求解

4.1 问题一：基于 LSTM 的气候变化时空数据关键信息提取及趋势分析模型

4.1.1 数据来源及预处理

4.1.1.1 数据来源

对于加拿大历史时空数据进行采集，本文选用的气象时空数据来源于 63 个加拿大城市，数据统计项分为平均温度和最高温度以及最低温度，由于加拿大幅员辽阔并且领土分布在北纬 41° — 83° 、西经 52° ~ 141° 之间，地形气候类型较多，基于我们的“局部映射整体”思想，为使得我们的数据能够更加好的反映加拿大全国温度的时空变化趋势，我们从加拿大全国所有城市中选取 64 个城市，城市测点分布如图，城市的选取原则如下：

- A 测点应覆盖加拿大全国的各类气候类型地区；
 - B 测点应覆盖全国的不同海拔高度地区；
 - C 测点的布置应该覆盖全国沿海与内陆地区；
 - D 测点应该覆盖全国各省；
- 海洋的时空数据来源于题目所给附件。

4.1.1.2 数据的预处理的方式

对于各站点气象数据均经过严格的质量控制和均一性检验^[3]，对个别缺失数据采用相邻站点线性回归方法进行插补^[5]，保证经过处理修正后的气象数据具有很好的连续性，然后基于 RCLimDex 软件对数据异常值与错误值进行过滤筛选，以满足分析要求。

基于加拿大的 64 个站点，每个待检站点的邻近站必须满足以下条件：

- A 邻近站点与待检测站点之间的水平距离在 350KM 以内；
- B 当待检站海拔高度在 2500m 及以下的时候，邻近站与待检站的高度差应该小于 200m；
- C 当待检站海拔高度在 2500m 及以上的时候，邻近站与待检站的高度差应该小于 500m；

均一性检验的数学模型如下：

将满足基本条件的邻近站及待检站的逐年气温序列进行差分处理，得到差分时间序列 d_T/d_i 上升或下，如式 1：

$$(d_T/d_i)_i = T_i - T_{i-1} \quad (1)$$

其中，T 代表气温，i 表示年份。

计算待检站 d_T/d_i 序列与临近站 d_T/d_i 序列的相关系数 rk1，针对每个待检站找出与其相关较高的 4~5 个站（以下简称高相关站）。

$$r_{k1} = \frac{\sum_{i=1}^m \left(\frac{dT}{dt} \right)_{ki} \left(\frac{dT}{dt} \right)_{li} - n \overline{\left(\frac{dT}{dt} \right)_k} \overline{\left(\frac{dT}{dt} \right)_l}}{\sqrt{\sum_{i=1}^m \left(\frac{dT}{dt} \right)_{ki}^2 - n \left[\overline{\left(\frac{dT}{dt} \right)_k} \right]^2} \sqrt{\sum_{i=1}^m \left(\frac{dT}{dt} \right)_{li}^2 - n \left[\overline{\left(\frac{dT}{dt} \right)_l} \right]^2}} \quad (2)$$

其中，k 代表待检站的差分序列，l 代表临近站差分序列，n 表示时间序列长度，i 表示年份。

将每个待检站对应高相关站的年尺度气温序列分别求其气候平均（1980 年—2006 年）序列及距平序列，同样方法计算月尺度气候平均序列及距平序列。

将（2）中所得年尺度气候值序列和距平序列分别求得算术平均序列并：求和，即得到待检站的参考序列，同样计算月尺度参考序列。

同时为了保证所建立的参考序列的合理性，需先对求得的参考序列进行验证：对年参考序列及待检站利用公式（1）分别进行差分处理，并利用公式(2)将两个差分序列进行相关系数的运算，若系数 ≥ 0.8 ，且通过 PMFT 检验为均一的，那么此序列便可作为该待检站的参考序列，否则参考序列需重建：对月尺度的参考序列验证则相对简单些，只需计算所求参考序列与待检站月尺度序列的相关系数，若系数 ≥ 0.8 ，并通过 PMFT 检验为均一即可。

利用 matlab 绘制全球海洋时空云图，部分云图如图 1，图中的颜色代表在一个特定时间上海洋温度的高低。

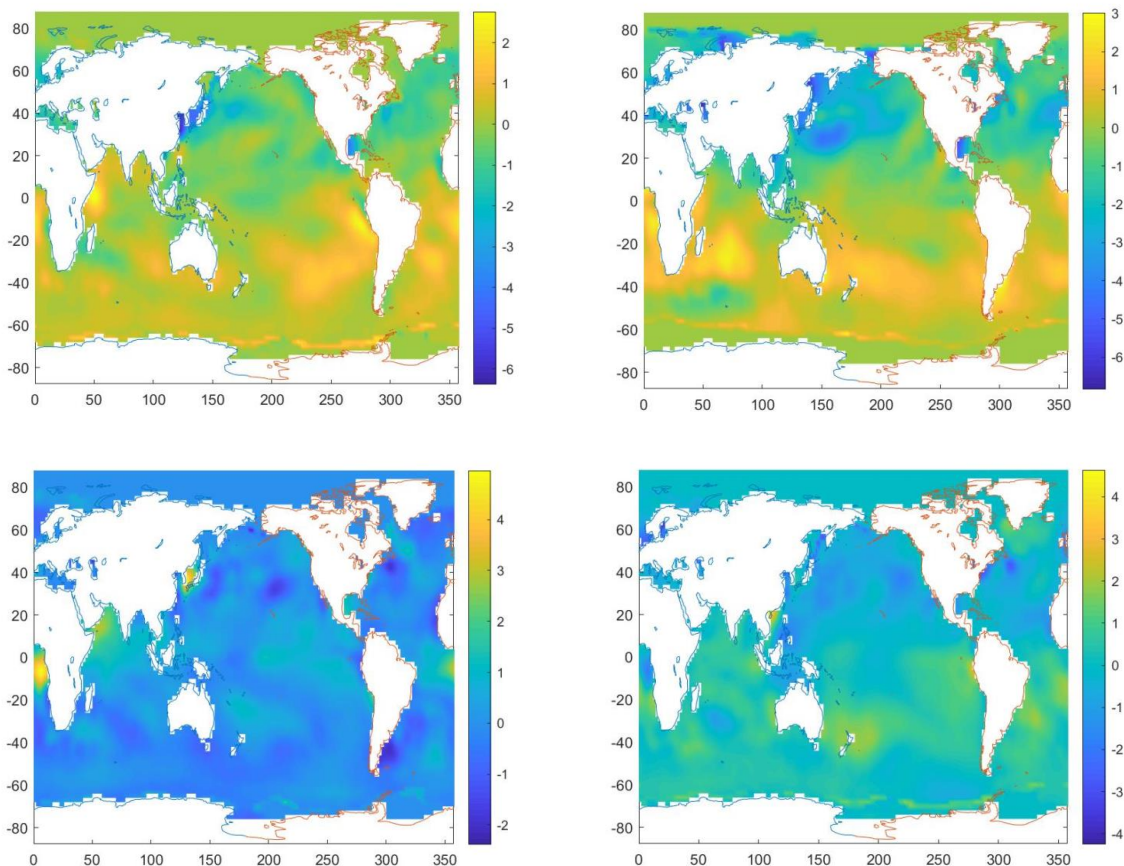


图 1 全球海洋时空云图（部分）

为了对全球海洋温度的变化趋势有更好地把控，分别选取两年、五年、十年、二十年作为时间间隔绘制海洋温度差值云图，现给出部分云图，全部云图见附件。

全球海洋每两年温度差值云图示例如下：

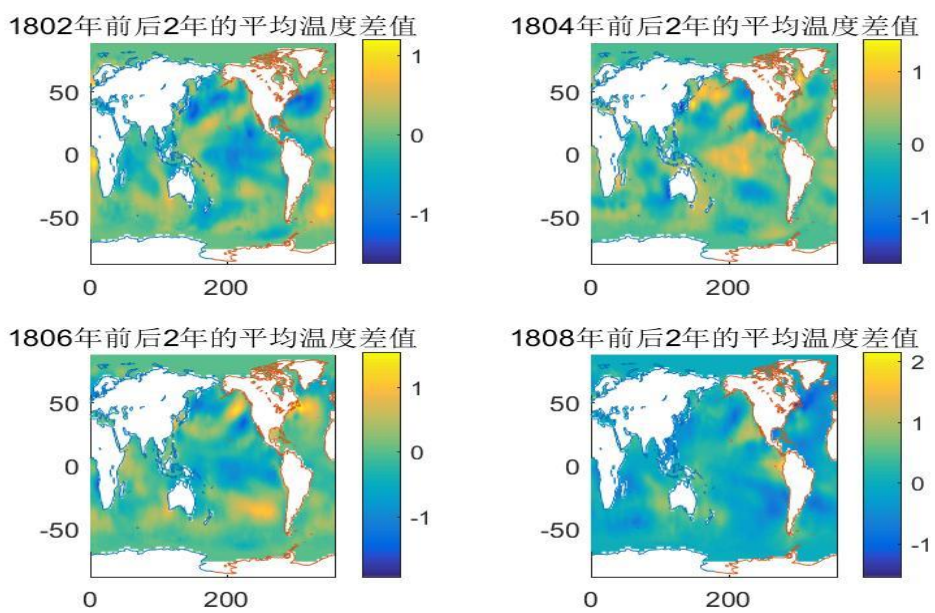


图 2 全球海洋差值时空云图（部分）

全球海洋每五年温度差值云图示例如下：

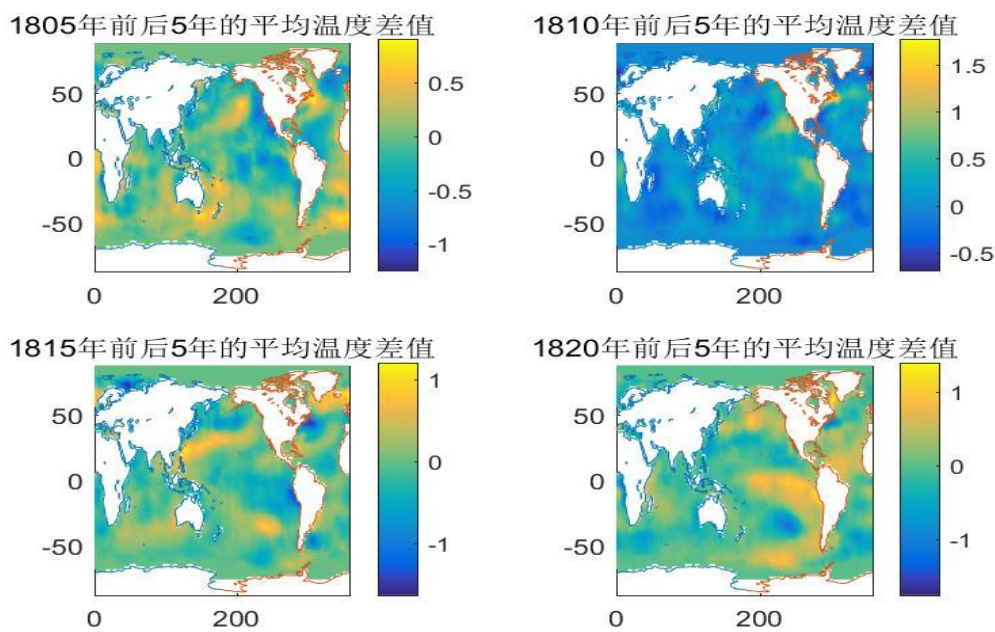


图 3 全球海洋差值时空云图（部分）

全球海洋每十年温度差值云图示例如下：

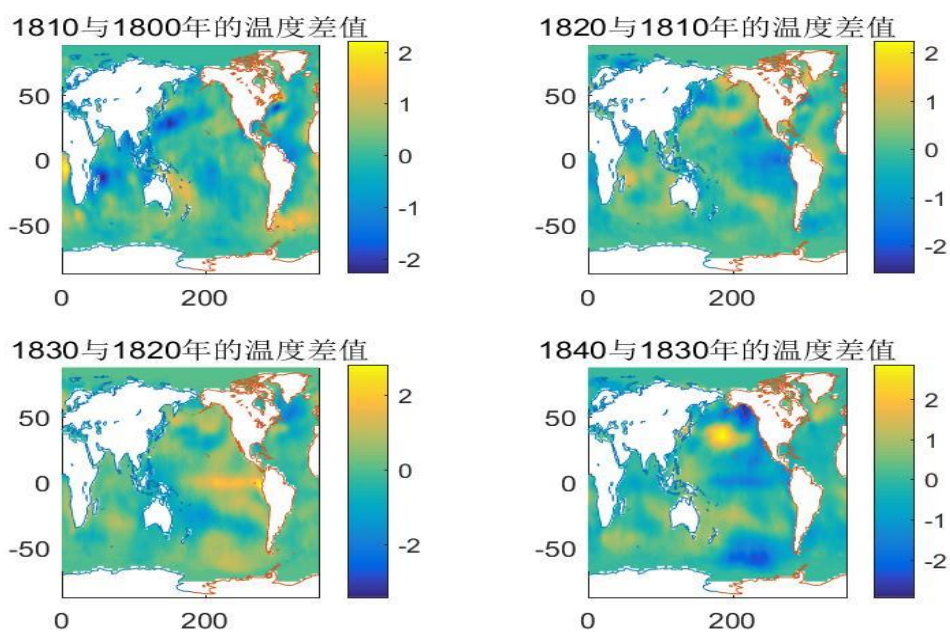


图 4 全球海洋差值时空云图（部分）

全球海洋每二十年温度差值云图示例如下：

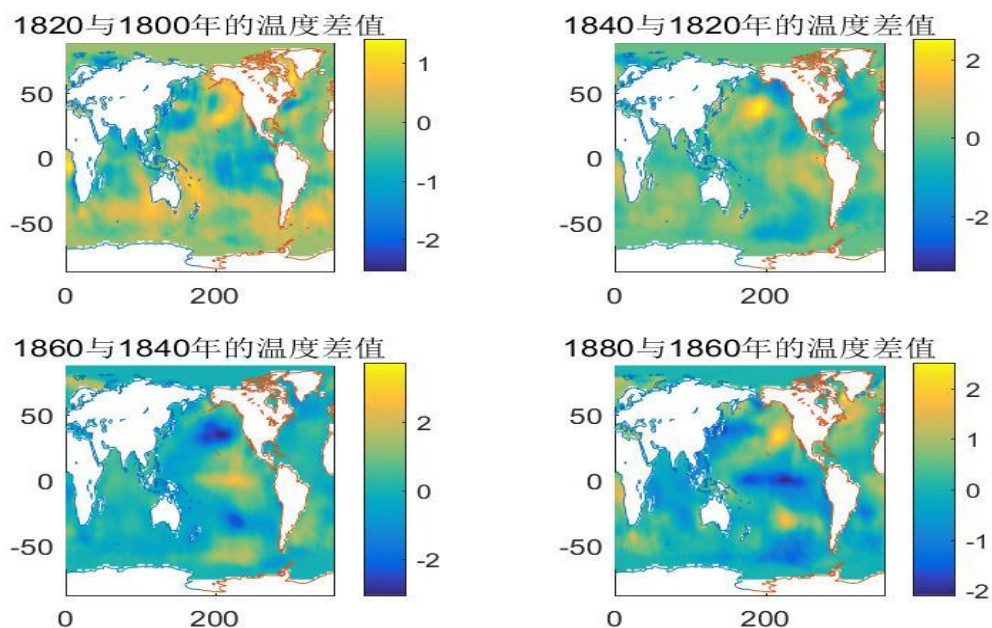


图 5 全球海洋差值时空云图（部分）

4.1.2 时间空间输入序列数据集

加拿大温度数据和全球海洋数据集均与时间空间强相关，以上数据集输入形式如下： $T[(u, v), t]$ 。T 表示全球海洋或加拿大温度值；u 为经度；v 表示纬度；t 表示时间。

4.1.3 基于 M-K 的非参数统计检验分析

Mann-Kendall 方法是一种非参数统计检验方法^[6]，具有样本不需要遵从一定分布、结果不受少数异常值干扰、计算过程方便且适用于类型变量和顺序变量等优点，适用性较强，受到世界气象组织高度推荐。该方法不但可以检验气象要素等时间序列的变化趋势，还可以检验气象要素时间序列是否发生了突变。利用 M-K 检验可以对上述数据进行统计检验。

(1) Mann-Kendall 趋势检验

对于一个气温时间序列 $X_i = (x_1, x_2, \dots, x_n)$ ，Mann-Kendall 趋势检验计算公式如下：

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (3)$$

式中： S 为 Mann-Kendall 趋势检验的统计量； x_i 表示气温时间序列的第 j 个 $2 \leq j \leq n$ 数据值； n 为气温时间序列长度；而 sgn 则为符号函数，其定义如下：

$$\text{sgn}(\theta) = \begin{cases} 1, \theta > 0 \\ 0, \theta = 0 \\ -1, \theta < 0 \end{cases} \quad (4)$$

Mann 和 Kendall 证明，当 $n \geq 8$ ，统计量 S 近似于正态分布，其均值为 0，方差为：

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^n t_i(i-1)(2i+5)}{18} \quad (5)$$

式中： t_i 是第 i 组的数据点的数目。

标准化统计量 Z ，可按照如下公式计算：

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}}, S > 0 \\ 0, S = 0 \\ \frac{S+1}{\sqrt{\text{Var}(S)}}, S < 0 \end{cases} \quad (6)$$

即服从标准正态分布。Mann-Kendall 趋势检验主要是依据统计量 Z 所处的置信区间范围来判断气温时间序列变化是否显著，当 $|Z| > Z_{1-\alpha/2}$ 时（ α 为显著性检验水平），表明气温时间序列在 α 显著性检验水平下变化显著。一般情况下， α 显著性检验水平取值为 0.05 或 0.01，其对应的 $|Z|$ 值分别为 1.96 和 2.58。

另衡量气温变化趋势大小的指标为倾斜度 β ，其计算公式为：

$$\beta = \text{Median}\left(\frac{x_i - x_j}{i - j}\right) \quad (7)$$

式中： $1 < j < i < n$ ， n 为气温时间序列的长度； β 值为正，表示气温呈“上升趋势”，

而 β 值为负，则表示气温呈“下降趋势”， $|\beta|$ 值越大说明气温时间序列的上升或下降的变化幅度就越大。

(2) Mann-Kendall 突变检验

Mann-Kendall 突变检验方法^[7]，常用与识别气象和水文领域中的突变现象，可明确指出突变区域以及突变开始的时间点。对于样本量为 n 的气温时间序列 $X_i = (x_1, x_2, \dots, x_n)$ ，首先可构造一个如下的秩序列：

$$S_k = \sum_i^k r_i \quad (k = 2, 3, \dots, n) \quad (8)$$

式中：

$$r_i = \begin{cases} +1, & x_i > x_j \\ 0, & x_i < x_j \end{cases} \quad (j = 1, 2, \dots, i)$$

秩序列 S_k 是第 i 个时刻气温数值大于第 j 个时刻数值时的次数累计数。

在假定气温时间序列为随机的情况下，可定义统计量 UF_k 为：

$$UF_k = \frac{[S_k - E(S_k)]}{\sqrt{Var(S_k)}}, (k = 1, 2, \dots, n) \quad (9)$$

式中： $UF_1 = 0$, $E(S_k)$ 和 $Var(S_k)$ 分别表示秩序列 S_k 的均值和方差。当气温时间序列

$X_i = (x_1, x_2, \dots, x_n)$ 相互独立，且呈现相同的连续分布， $E(S_k)$ 和 $Var(S_k)$ 可分别由公式计算出：

$$\begin{aligned} E(S_k) &= \frac{k(k-1)}{4}, (2 \leq k \leq n) \\ Var(S_k) &= \frac{k(k-1)(2k+5)}{72}, (2 \leq k \leq n) \end{aligned} \quad (10)$$

式中： UF_k 为标准正态分布，它是按气温时间序列 X_i 的顺序 (x_1, x_2, \dots, x_n) 计算出的统计量

序列，在给定显著性检验水平 α 的情况下，查正态分布表，若 $UF_k > U_\alpha$ ，则表明气温时间

序列 $X_i = (x_1, x_2, \dots, x_n)$ 存在明显的变化趋势。然后，在按气温时间序列 X_i 的逆序列

$(x_n, x_{n-1}, \dots, x_1)$ ，重复上述计算过程，同时使 $UB_k = -UF_k (k = n, n-1, \dots, 1), UB_1 = 0$ 。

一般情况下，取显著性水平 $\alpha = 0.05$ ，则临界值 $U_{0.05} = \pm 1.96$ 将 UF_k 和 UB_k 两个统计量序列曲线与 ± 1.96 两条直线共同绘制在一张图上。若 UF_k 和 UB_k 的值大于 0，则表明气温时

间序列 $X_i = (x_1, x_2, \dots, x_n)$ 呈上升趋势, $X_i = (x_1, x_2, \dots, x_n)$ 小于 0 则表明气温时间序列呈下降趋势。当 UF_k 和 UB_k 的值超过临界直线时, 表明气温时间序列 $X_i = (x_1, x_2, \dots, x_n)$ 上升或下降趋势显著, 超过临界直线的范围即为气温时间序列发生突变的时间区域。如果 UF_k 和 UB_k 在临界直线之间出现交点, 则该点可被确定为气温时间序列的突变点, 对应的时刻为气温突变开始的时间。

4.1.4 LSTM 气候时空趋势变化模型的建立

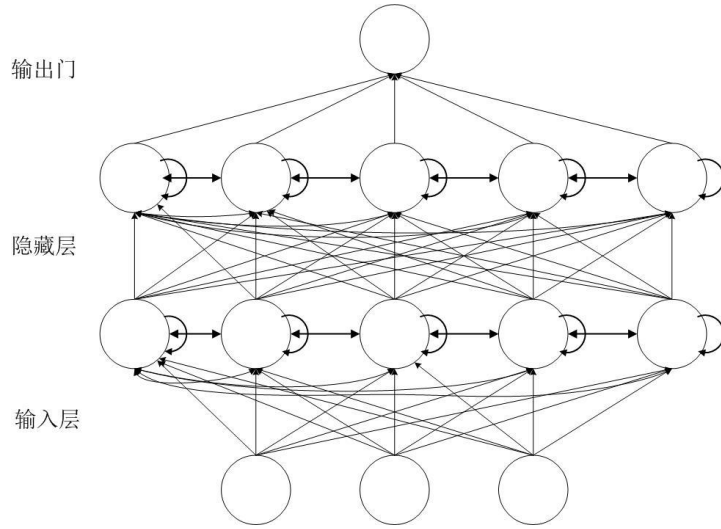


图 6 LSTM 模型结构

LSTM 模型用记忆块取代传统的隐含节点^[8], 使梯度在训练展开比较长时不消失或爆炸, 从而解决了传统神经网络模型无法处理长时间序列的难题。基于 KNN 的改进 LSTM 模型核心部分为记忆块, 由输入门和遗忘门以及输出门组成, 一个记忆块在 t 时刻的计算过程如下:

$$\begin{aligned}
 \tilde{c}_t &= \tanh(w_{xc}x_t + w_{ch}h_{t-1} + b_c) \\
 i_t &= \sigma(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci}c_{t-1} + b_i) \\
 f_t &= \sigma(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf}c_{t-1} + b_f) \\
 o_t &= \sigma(w_{xo}x_t + w_{ho}h_{t-1} + w_{co}c_{t-1} + b_o) \\
 c_t &= c_{t-1} \otimes f_t + i_t \otimes \tilde{c}_t \\
 h_t &= o_t \otimes \tanh c_t
 \end{aligned} \tag{10}$$

式中: \tilde{c}_t 表示 t 时刻记忆细胞的更新状态; i_t 、 f_t 、 o_t 、 c_t 、 h_t 分别表示输入门、遗忘门、输出门、记忆细胞以及隐藏层在 t 时刻的输出; x_t 表示 t 时刻的输入; h_{t-1} 、 c_{t-1} 分别表示隐藏层及记忆细胞在 $t-1$ 时刻的输出; w_{xc} 、 w_{ch} 分别表示记忆细胞与输入 x_t 以及隐藏层的权值矩阵;

w_{xi} 、 w_{hi} 、 w_{ci} 分别为输入门与 x_t 、隐藏层、记忆细胞的权值， w_{xf} 、 w_{hf} 、 w_{cf} 分别为遗忘门与 x_t 、输出层、记忆细胞的权值， w_{xo} 、 w_{ho} 、 w_{co} 分别为记忆细胞与 x_t 、输出层、记忆细胞的权值； \otimes 为点乘， σ 为 sigmoid 激活函数； b_c 、 b_i 、 b_f 、 b_o 为偏倚^[9]。

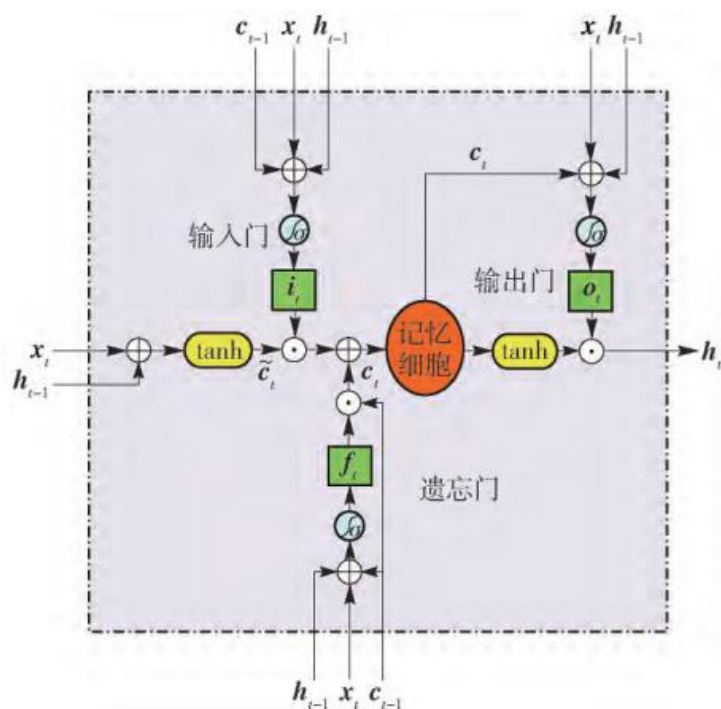


图 7 LSTM 记忆细胞结构示意图

4.1.5 模型的求解与结果

(1) 模型的求解

上述 LSTM 的气候变化时空数据关键信息提取及趋势分析模型求解过程如下：

Step1: 原始数据输入

输入收集的加拿大温度时空数据以及全球海洋时空数据

Step2: 数据集化

将原始数据进行输入时空数据集化作为程序自变量的输入：

$$T[(u, v)] \quad (11)$$

Step3: M-K 非参数统计

输出趋势检验指标如下：

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (12)$$

输出突变趋势检验检验如下：

$$UF_k = \frac{[S_k - E(S_k)]}{\sqrt{Var(S_k)}}, (k= 1, 2, \dots) \quad (13)$$

Step4: 数据集输入 LSTM 气候时空趋势变化模型:

时空趋势模型通过 Python 进行程序实现, 并且将 Step2 中的数据集输入程序。

Step5: 结果输出

结果输出项如下:

A:M-K 趋势检验指标^[10]; B:M-K 突变趋势检验指标; C:平均气温线性拟合图;

D:timeliness-tendency 值并绘制线性趋势空间云图; E:全球时空方差云图; F:分区海洋温度变化分析图 G:全球海温年变化云图

问题一解决全过程流程图如图 8 所示:

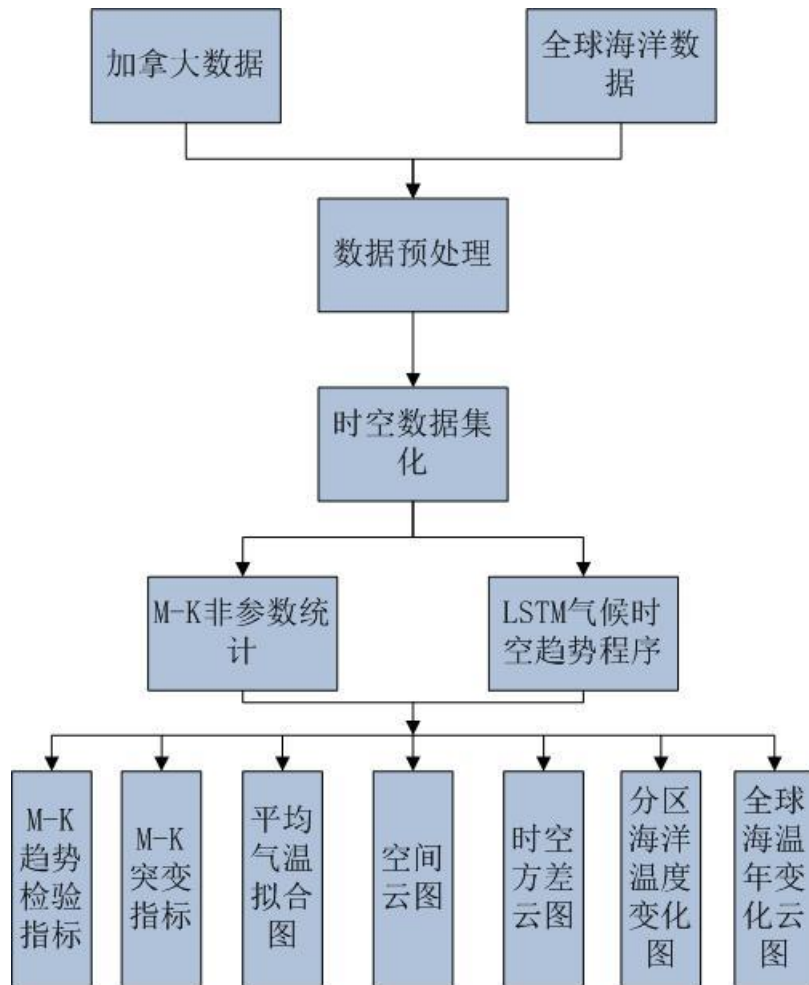


图 8 问题一解决全过程流程图

(2) 加拿大地区温度时空变化趋势结果

站点选取了加拿大的 150 个城市的气候数据, 由于不同站点之间收集到的气候数据的时间段不一定相同, 且由于年代久远, 收集到的某些区域某些时间段的气候数据会出现缺失的情况。为了使分析的数据的数量满足要求, 对个别缺失数据采用相邻站点线性回归方法进行插补, 保证经过处理修正后的气象数据具有很好的连续性, 然后基于 RCLimDex 软件对数据异常值与错误值进行过滤筛选, 以满足分析要求。即根据空间相近, 时间相同原则, 选取相近的坐标点并求其温度的平均值来作为缺失点的温度值。

由于各个站点收集的气候条件数据横跨的时间范围较广, 且可能收集的时间范围长短

参差不齐，或不同观测站点之间的时间观测范围没有交集，使得不同监测站点的气候数据对比困难，故本次数据在相邻站点线性回归的基础之上，在不同站点之间的时间范围尺度上寻找最大的公共时间尺度，找到适合不同站点的完整时间区段的数据，并进行筛选。

最终选用 63 个站点的统计值，包括 1840 年至 2006 年 26 年间每月的平均温度、最低温度、最高温度、平均降水量、最小降水量、最大降水量、降雪量等，详见附表。

(3) 加拿大全国年平均气温变化分析：

对 63 个站点的年平均温度求平均值，汇出 1980 年—2006 年的年平均温度折线图，并对其进行拟合，得到 26 年间温度变化的大致趋势，该趋势用 *timeliness-tendency*(TK) 指标来评价，即表示该拟合线段的斜率值，斜率越大，说明该地区在这 26 年间温度变化的，如图所示。

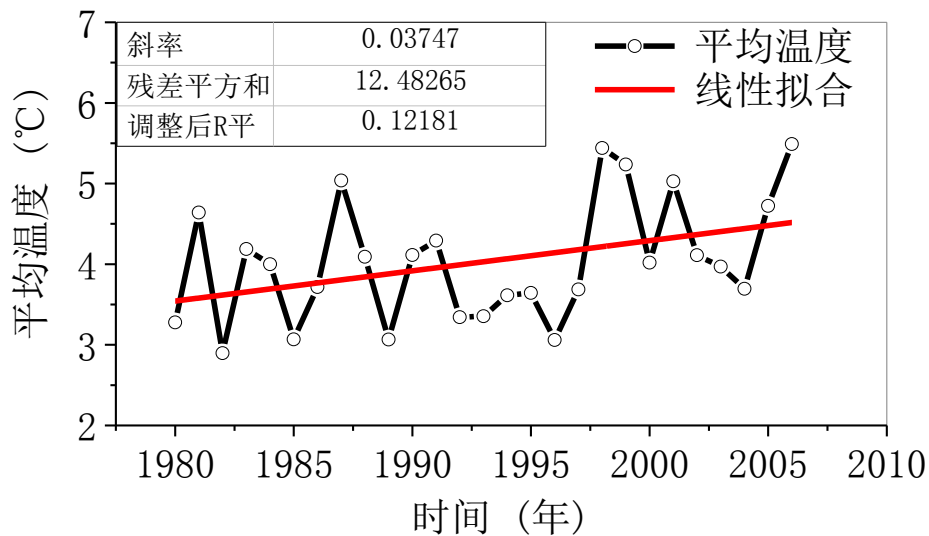


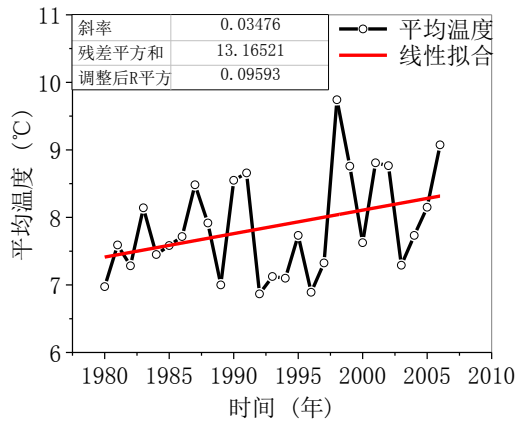
图 9 全国年平均气温变化图

由图可知，在统计的 26 年区间，年平均温度是不断波动的，但整体呈上升趋势；图中，用于评价温度变化趋势的 *timeliness-tendency* 值为正，也能基本判定温度上升的整体趋势。最低温度和最高温度差值在 2.5°C 左右，在 1994 年—1996 年间，加拿大的年平均气温较低。

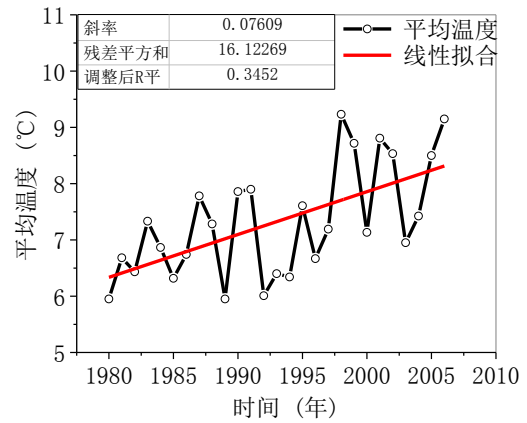
(4) 加拿大各地区观测站点年平均气温变化分析：

绘出加拿大 63 个站点在 26 年间的年平均气温变化图，并且按上述方法对其进行拟合，对其的整体趋势进行分析，全部数据图见附件。

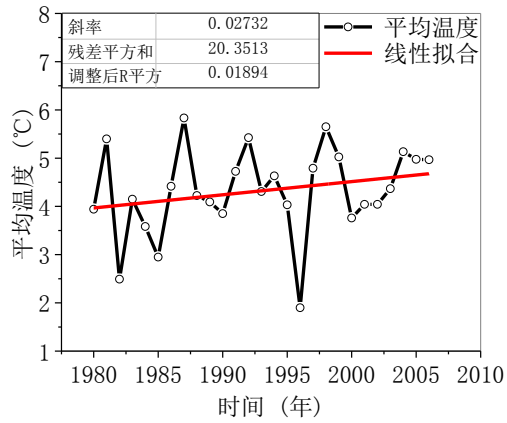
对 63 个站点的气候数据图根据经纬度来进行聚类分析，把相近经纬度的站点划分为一类，本次统计中一共分为 5 类。从每一类中随机挑选出一张线性趋势空间格局图，对每个站点的年平均气温的变化趋势进行分析，如图所示。



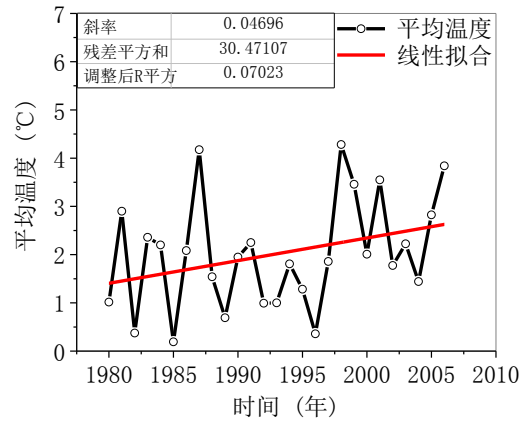
HAMILTON A1



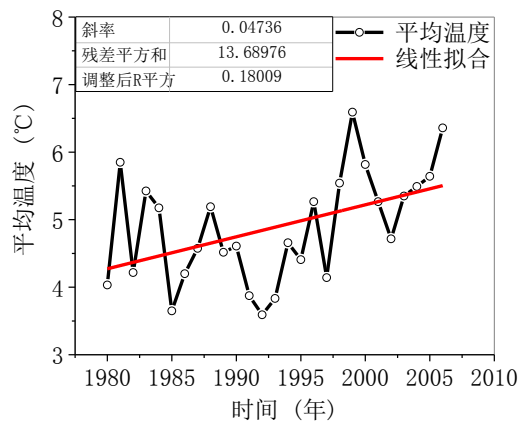
KINGSTON A1



PRINCE GEORGE A 1



SIOUX LOOKOUT



ST JOHN'S2

图 10 各站点年平均气温变化图

由图可知，各个站点的年平均气温变化趋势都是不断上升的，同样用 *timeliness-tendency* 来表征，按照聚类方法分类之后随机抽取的 5 个站点的温度变化趋势

与全国的年平均气温变化趋势是相符的。在 1995 年—2000 年，5 个站点的年平均气温都发生了较大的波动，说明该在这 5 年间，出现了极热或极寒天气。

把 63 个站点依据时间属性空间属性划分为 5 类“站点”，通过 matlab 计算出这 5 个“站点”的年平均温度，绘出站点在 26 年间的温度变化曲线。

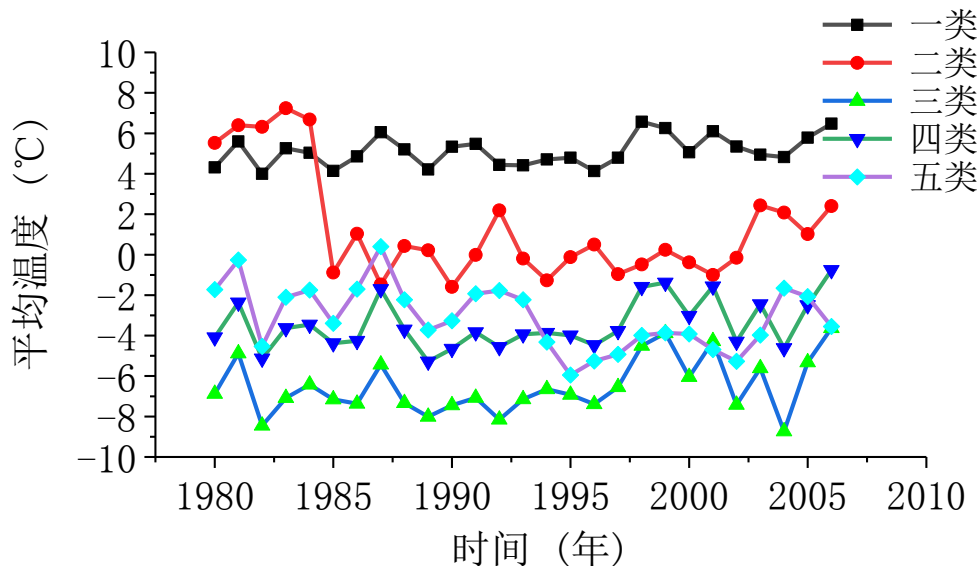


图 11 5 类“站点”年平均温度变化图

由图可知，在 1980 年—2006 年间，5 个“站点”的年平均温度的变化趋势总体相似。在 1981 年、1987 年、1991 年、1994 年、1999 年 2003 年出现波峰，在 1982 年、1985 年、1989 年出现波谷。1995 年前，第三、四、五类站点变化波动相似，第二类站点变化差别较大，1995 年之后，第二、三、四类站点波动趋于相似，第五类站点与前三类差别较大。第一类站点的温度与其他四类站点相比，波动趋势相似，但其波动幅度比较小，整体趋势较为平稳。在 26 年间，第二类站点的温度下降了 4°C 左右，第三、四类站点上升 2-3°C，第五类站点下降 2°C 左右，第一类站点变化较小。

(5) 加拿大各地区观测站点线性趋势空间格局分析：

根据前面各站点年平均气温变化图中拟合出的 *timeliness-tendency* 值，在空间上对其进行分析。*timeliness-tendency* 值的绝对值大小反映了该地区在 26 年间的年平均气温的变化速率，TK 值为正时，TK 越大，表明该地区的年平均气温上升得越快；TK 为负是，表明该地区在 26 年间的年平均气温呈下降趋势。把评价标准 TK 的变化范围划分为 8 个带，由冷色带向暖色逐渐增加，通过观察图中颜色的变化可以快速看出某地区 26 年间的温度速率增减情况。

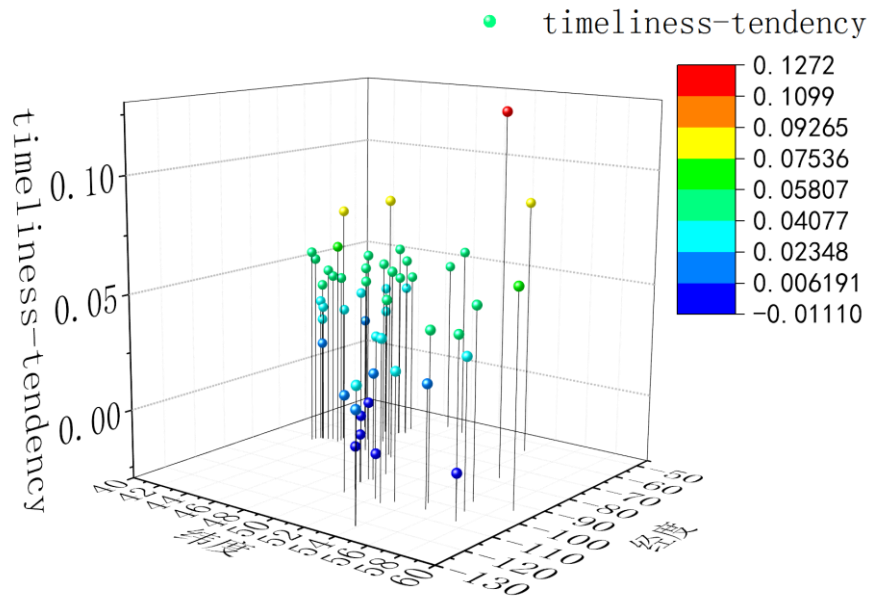


图 12 线形趋势空间格局图

由图可知，在统计的 63 个站点中，有 97%地区的评价指标 TK 值都为正^[12]，说明加拿大绝大部分地区的温度值在 26 年间都呈现了上升的趋势，但 80%站点的评价指标 TK 值都控制在 0—0.04 之间。图中观察到在 KUUIJUAPIK A 1 站点处(纬度 55.28, 经度-77.75)的 TK 最大，为 0.127，说明该地区在 26 年间的温度上升是最快的，此处的气候环境变化很大。同时，可以观测出低纬度地区的温度上升趋势普遍高于低纬度地区。

(6) 加拿大各地区观测站点年平均气温方差变化分析：

为了统计加拿大 63 个测点的年平均温度离散情况，对每个站点在 26 年间的温度变化情况求其方差，用 timeliness-variance (TK) 来表现，TK 值越大，表明该地区在 26 年间的温度变化范围越大。

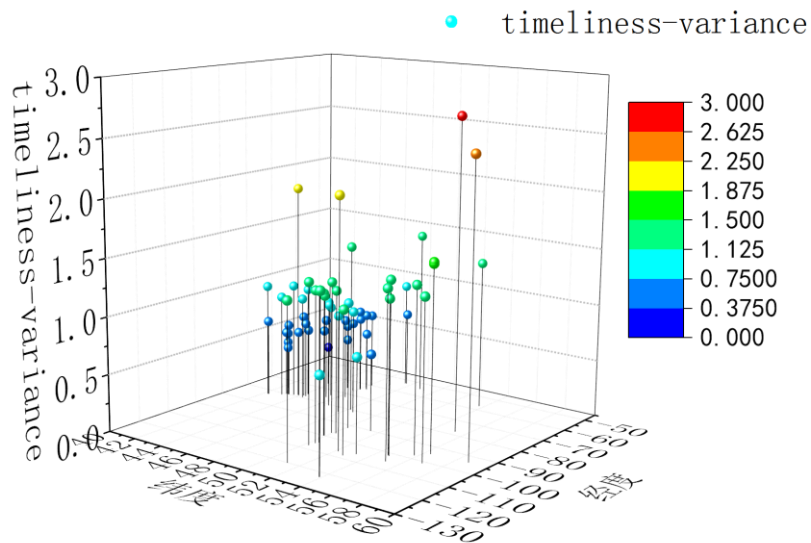


图 13 各站点年平均气温方差变化图

由图可知,63个站点的温度离散程度大致相近,都在0.5至1.5之间波动。在纬度55.28,经度-77.75和纬度58.74,经度-94.07站点处的方差值最大,分别为2.7和2.5。说明这两个站点处在26年间的温度波动范围最大。

(7) 加拿大各地区观测站点年平均温度变化分析:

把收集到的63个站点在26年间的年平均温度反映到空间上来,根据各个站点的经纬度建立三维空间温度变化曲线,可以直观的观察出空间上不同站点的温度情况。

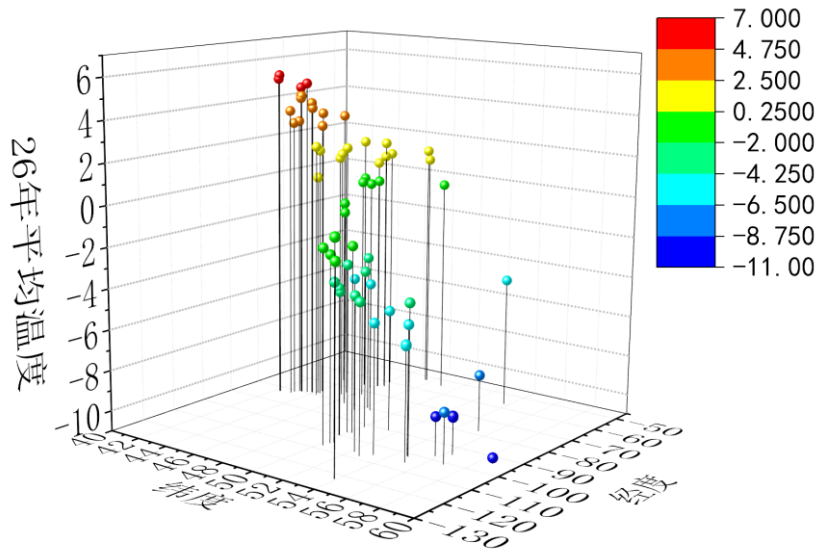


图 14 26 年平均温度空间分布图

由图可知,随着纬度的升高,平均温度基本呈下降,最大温差在15℃左右,但在同一纬度范围内,不同经度地区的26年平均温度有所波动,是由于加拿大的地域辽阔,不同地区的地貌特征有所差异,各地的海拔高度也不尽相同,加上各地的工业发展程度以及人口密度的不同,会造成温室气体排放量的不同,对温度影响情况也不同。

(8) 加拿大不同纬度地区在26年间的年平均温度分析:

将63个站点的年平均气温求平均得到1980-2006年的各站点的历史平均气温,按照纬度作出折线图,可以反映不同纬度带温度的大致分布情况,如图所示。

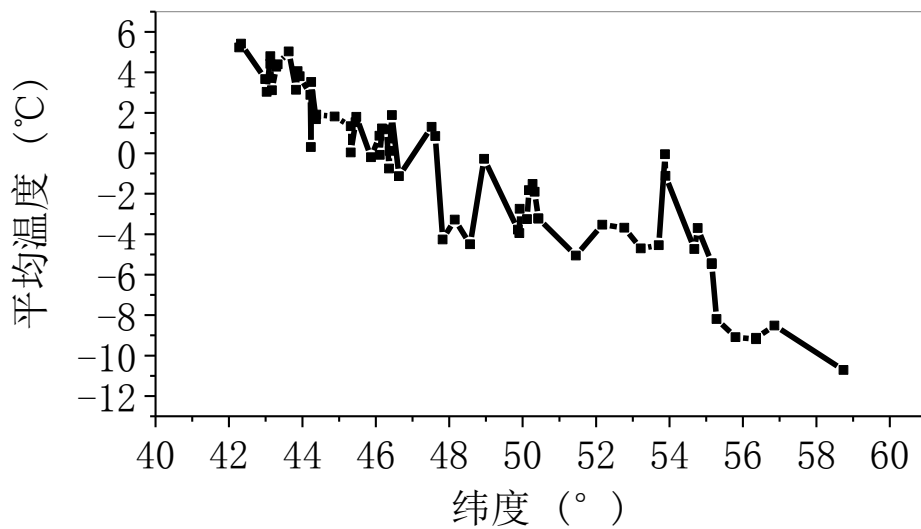


图 15 不同纬度带的温度分布情况图

由图可知，随着纬度的增加，温度大体呈现出下降的趋势。在北纬 54 度附近出现了一个波峰点，是由于该地区工业发达，温室气体排放量多，造成局部温度高的现象，表明温室气体对局部的大气系统影响明显。

(8) 海洋表面温度历史数据规律结论

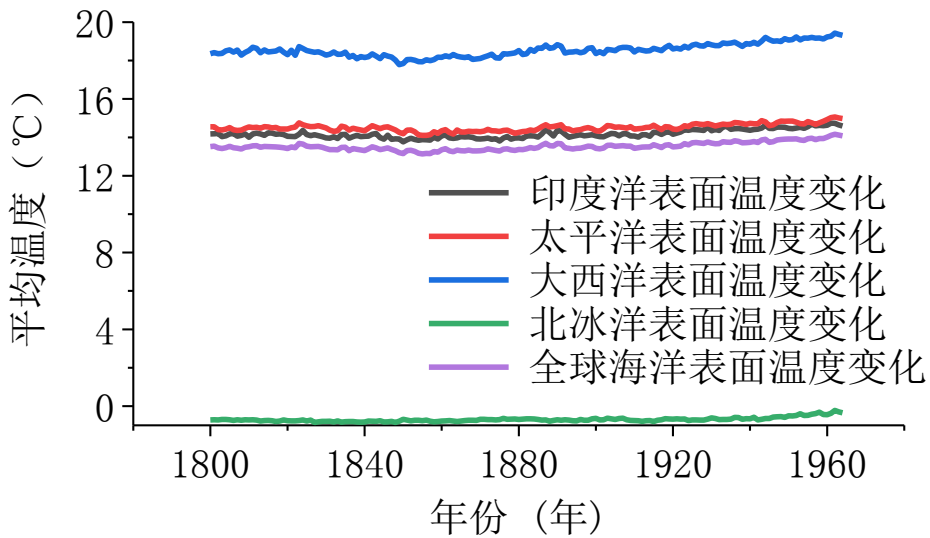


图 16 海洋温度变化曲线

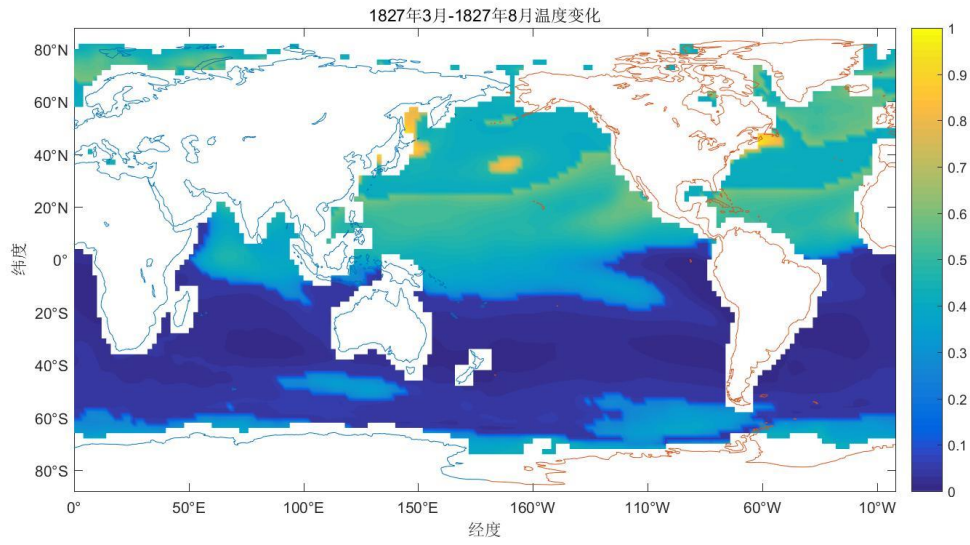


图 17 1827 年 3-8 月全球海温变化

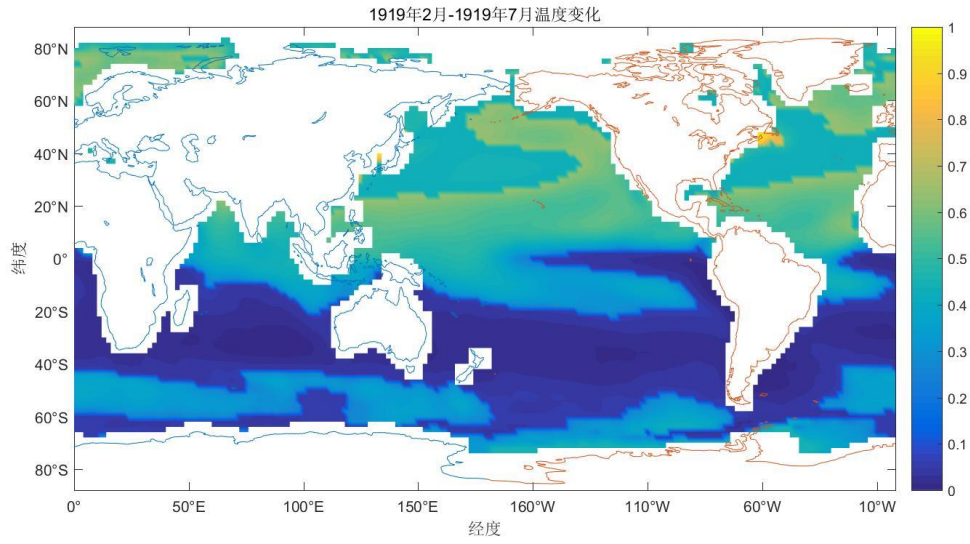


图 18 1919 年 2-7 月全球海温变化

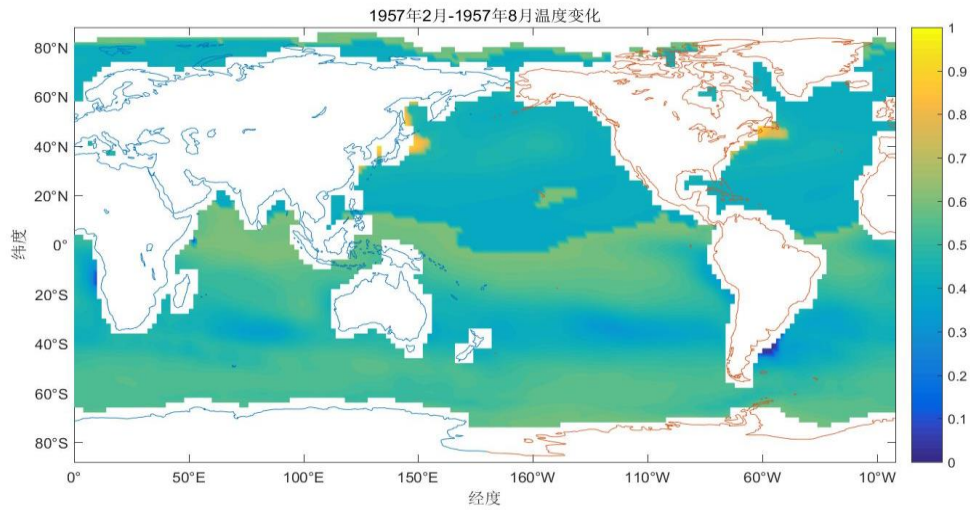


图 19 1957 年 2—8 月全球海温变化

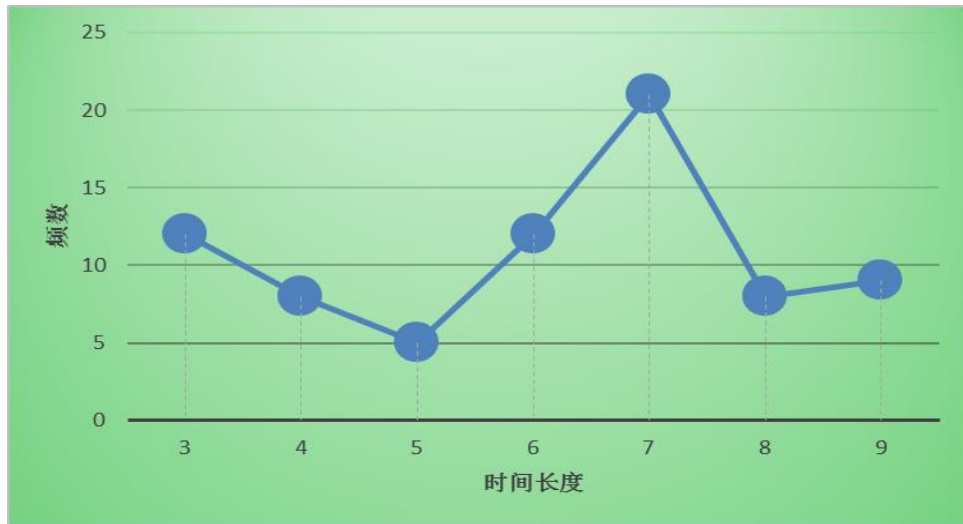


图 20 全球温度上升周期频率表

通过对全球每年海洋表面温度上升的月份及周期进行统计可知，全球海洋表面温度上升大多集中在 3-8 月份，上升时间大多集中于 7 个月左右。由于存在厄尔尼诺的现象以及其他的极端天气情况，导致全球海洋表面温度上升出现的年份、持续的时间不同于一年中海洋表面温度正常上升的情况。

从图上可看出，对全球海洋表面温度时间分布特征上的研究，随着年份的增加，在每年的 2 月到 8 月之间，海洋温度逐年上升。从时间序列 1800-1964 年上来看，北冰洋温度常年处于 0℃ 一下，是由于受到地理位置的影响所导致的特殊性。但从其他海洋，如太平洋的颜色变化趋势可以明显看出温度在一步一步的上升。从整体上来看，全球海洋的温度是在缓慢增长。由于海洋的升温，会加剧温室效应的加剧，并会提高大气中二氧化碳的含量，逐步导致全球变暖。

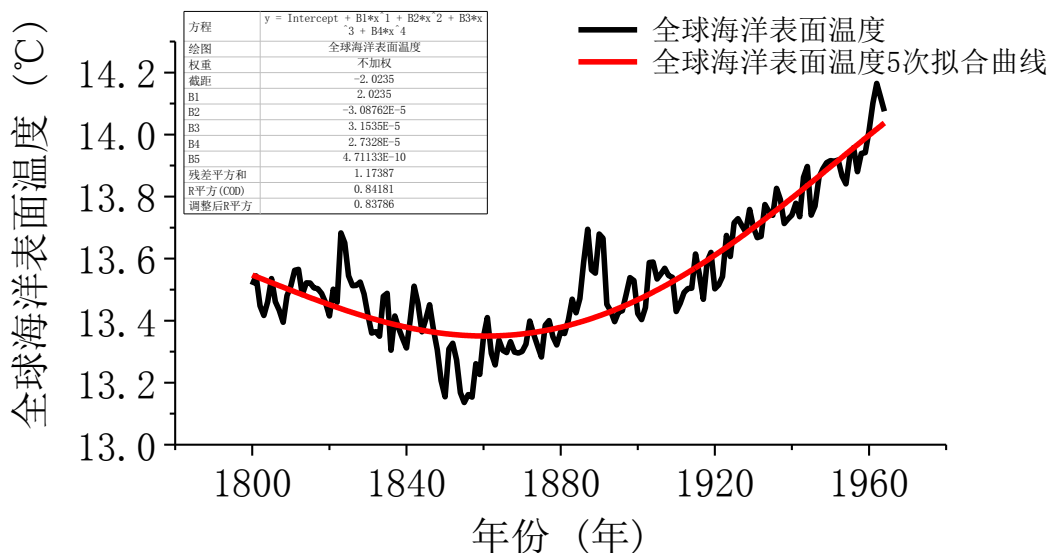


图 21 全球海洋表面温度变化曲线

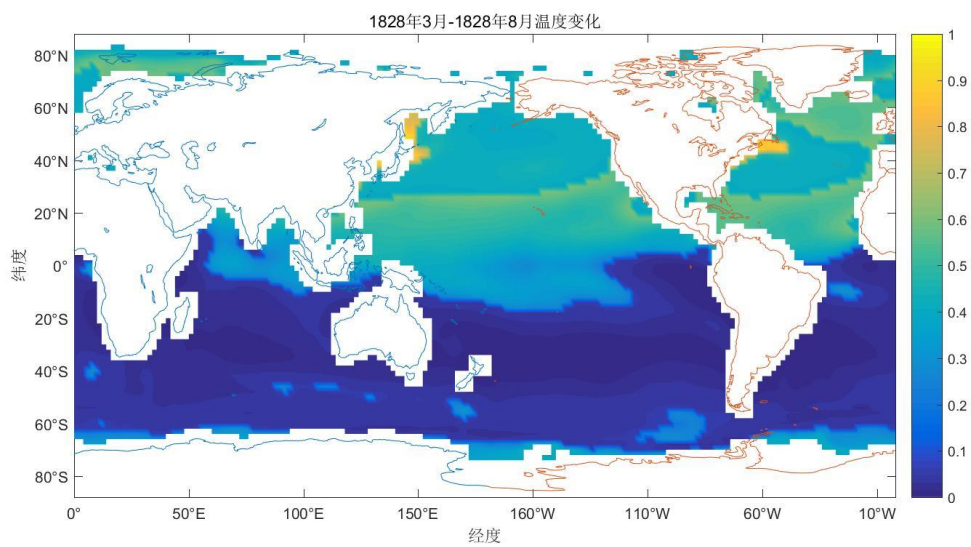


图 22 1828 年厄尔尼诺现象图

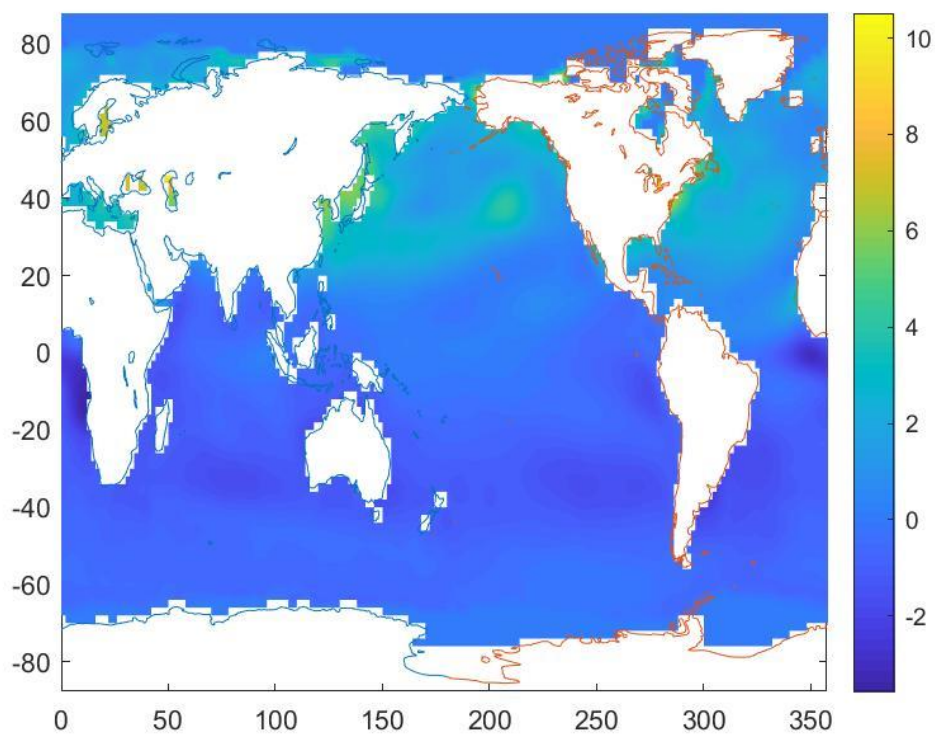


图 23 1891 年厄尔尼诺现象图

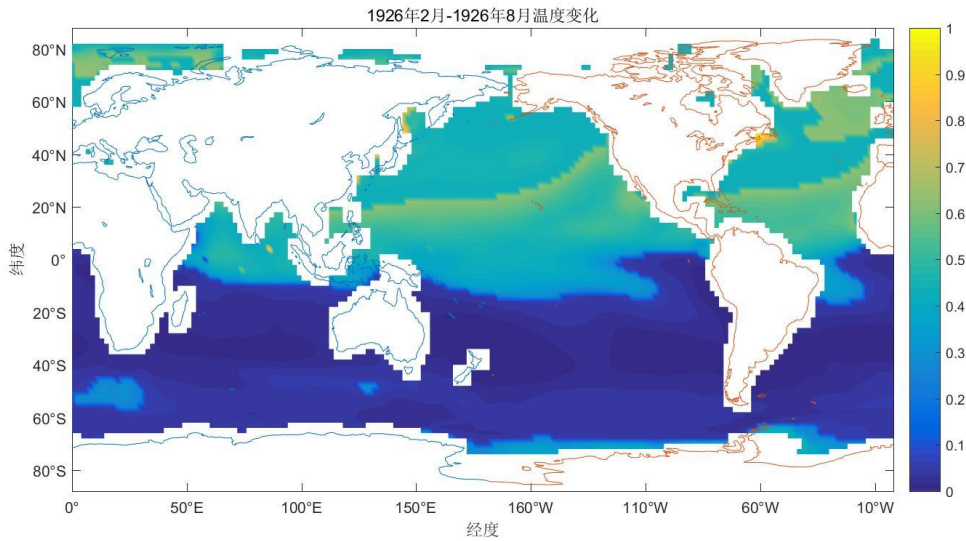


图 24 1926 年厄尔尼诺现象图

由图可看出，对空间分布特征进行分析，南、北半球大洋表面水温的变化相比，北半球的变幅大，这与盛行风的年变化有关，冬季来自大陆的冷空气，大大降低了海面温度；南半球的对应海域，由于洋面广阔以及经线方向洋流不像北半球那么强，所以变化幅度比较小。从图中也可以看出太平洋在全球和局部气候中都距有极为重要的作用，这是由于其特定的地理位置和活跃的西边界流系统，汇聚着来自不同海域的多种水团，是赤道太平洋与亚热带太平洋之间进行物质、热量以及盐量交换的主要区域。对产生厄尔尼诺-南方涛动（ENSO）事件的发生产生作用^[15]，南向的西边界流还为从太平洋流向印度洋的印尼贯穿流提供水源，从而对全球海洋的热盐环流也起到作用。

对图中全球海洋温度变化进行 5 次拟合，拟合曲线基本满足全球海洋表面温度变化规律，而且在 1828 年、1891 年、1926 年左右有着明显的突变，查资料可知，在这几年中均发生了 ENSO 事件。历史上，1858 年后，由于人类活动所带来的产业发展、化石燃料的高排放、森林砍伐等一系列活动所产生的温室气体逐步使海洋温度持续上升。并且海洋表面温度也以 0.2°C/10 年的速度持续增长。

由上述可知，1800 至 1964 年以来，虽然全球海洋表面温度局部变化是波动的，但全球海洋及各大洋整体相同的表面温度上升趋势表明温室效应引起的全球变暖趋势正逐步增强。

4.2 问题二：基于多元高次多项式非线性回归的复杂气候模型简化气候预测模型

4.2.1 相关因素影响机理分析及权重估计

(1) 机理分析

气候大气物理特征的长期平衡状态，与天气不同，气候具有稳定性，气候的这种稳定性可以表征为稳定不变或者以某种规律稳定变化，为研究复杂气候模型的“稳定模态特征”，需对引起气候变化的各类影响因素进行分析以及归类。

气候状态具有多种表征，以“SEA LEVEL—TEMPERATURE—FALL”作为预测目标，以自然因素和人为因素作为影响气候状态的三级指标，以太阳辐射等 6 个因素作为二级指标，以海洋温度等 18 个一级指标作为输入观测变量。

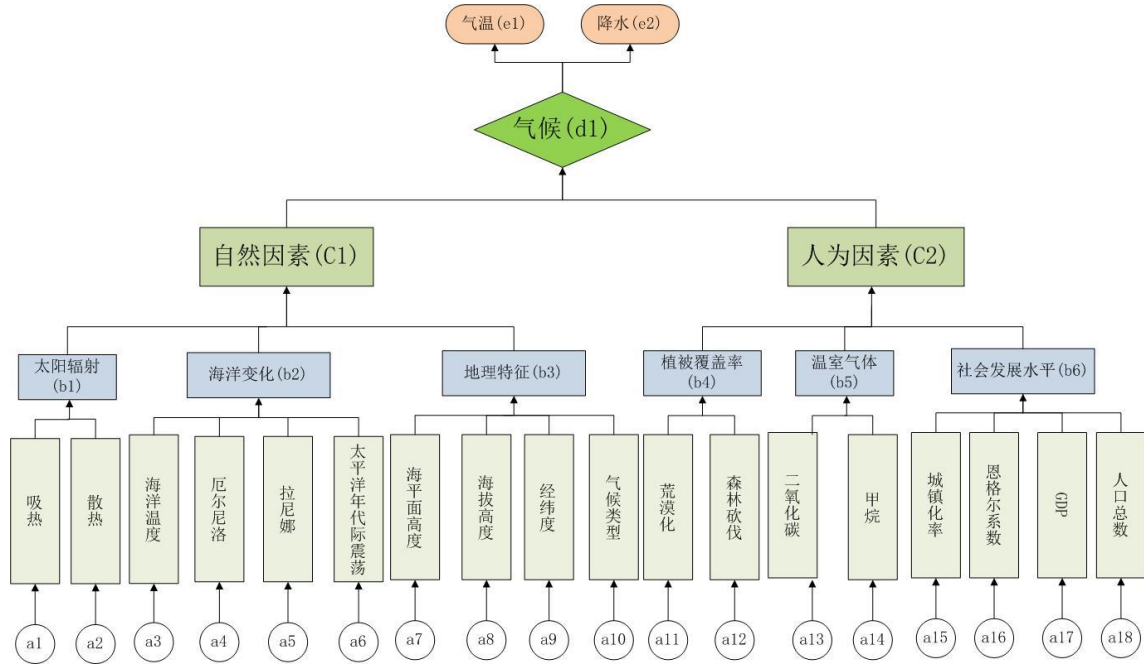


图 25 相关因素机理分析图

(2) 以 STF 为目标的多因素路径分析

气候特征的 STF 指标直接受到二三级指标的控制，实际上述一二级指标对于气候特征的影响权重是不一样的，需要考虑各类数据对于温度的影响程度大小，本路径分析基于如下带约束条件的最小二乘回归算法进行。

考虑到多因子模型中各个权值的实际意义，在训练权重时，需要加入约束条件。本文采用带约束条件的最小二乘回归算法^[17]，对上述多因子模型进行求解。对于 m 个样本来说，可以用如下线性方程组来表示：

$$\begin{cases} \beta_1\gamma_{11} + \beta_2\gamma_{12} + \beta_3\gamma_{13} = H_1 \\ \beta_1\gamma_{21} + \beta_2\gamma_{22} + \beta_3\gamma_{23} = H_2 \\ \beta_1\gamma_{31} + \beta_2\gamma_{32} + \beta_3\gamma_{33} = H_3 \\ \dots \\ \beta_1\gamma_{m1} + \beta_2\gamma_{m2} + \beta_3\gamma_{m3} = H_m \end{cases} \quad (12)$$

将样矩阵本记为 γ ，参数矩阵记为向量 β ，真实值记为向量 H ，上述线性方程用矩阵表示为：

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_{m1} & \gamma_{m2} & \gamma_{m3} & \gamma_{m4} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} H_1 \\ H_2 \\ \dots \\ H_m \end{pmatrix} \quad (13)$$

利用拉格朗日 (Lagrange) 乘子法，加入约束条件可得：

$$\varphi(\beta, \lambda, \mu) = \sum_1^m |H_i - \beta_2 \gamma_{i2} - \beta_3 \gamma_{i3}|^2 + \lambda \left(\sum_i^3 \beta_i - 1 \right) + \frac{c}{2} \left(\sum_i^3 \beta_i - 1 \right)^2 + \frac{1}{2c} \sum_j^3 \left\{ \left[\min(0, \mu_j + c \beta_j) \right]^2 - \mu_j^2 \right\}$$

式中： λ, μ 为拉格朗日乘子， c 为罚因子。进一步简化得到乘子迭代公式：

$$\lambda_j^{k+1} + \mu_j^{k+1} = \lambda_j^k + c \left(\sum_i^3 \beta_i - 1 \right) + \min(0, \mu_j^k + c \beta_j) \quad (14)$$

迭代计算结束准则为：

$$\left(\sum_i^3 \beta_i - 1 \right)^2 + \sum_i^3 \left[\min\left(\beta_j, -\frac{\mu_j}{c}\right) \right]^2 < \varepsilon^2 \quad (15)$$

ε 为计算精度，所有问题都归结求估计参数 $\beta_1, \beta_2, \beta_3$ ，误差达到最小值时就是回归方程的最优解。

基于上述以 STF 为目标的的多因子路径分析原理，利用 Python 进行程序实现求出各类影响因素的权重大小，略去各影响权重高阶微量得到如下简化后的影响机理图。

```
print(examDf[examDf.isnull()==True].count())
print(examDf.corr())
sns.pairplot(data, x_vars=['X6-1', 'X6-2', 'X6-3'], y_vars='y', height=7, aspect=0.8, kind='reg')
plt.show()
from warnings import simplefilter
simplefilter(action='ignore', category=FutureWarning)
X_train, X_test, Y_train, Y_test = train_test_split(examDf.iloc[:, :3], examDf.y, train_size=0.75)
print("自变量---源数据:", examDf.iloc[:, :3].shape, "; 训练集:", X_train.shape, "; 测试集:", X_test)
print("因变量---源数据:", examDf.y.shape, "; 训练集:", Y_train.shape, "; 测试集:", Y_test.shape)
model = LinearRegression()
# 调用线性规划包
model.fit(X_train, Y_train)
# 线性回归训练
a = model.intercept_ # 截距
b = model.coef_ # 回归系数
print("拟合参数:截距", a, ", 回归系数:", b)
print("最佳拟合线: Y = ", round(a, 3), "+", round(b[0], 3), "* (X6-1) + ", round(b[1], 3), "* (X6-2) +")
# 显示线性方程，并限制参数的小数位为两位
Y_pred = model.predict(X_test)
# 对测试集数据，用predict函数预测
```

图 26 多因子路径分析代码

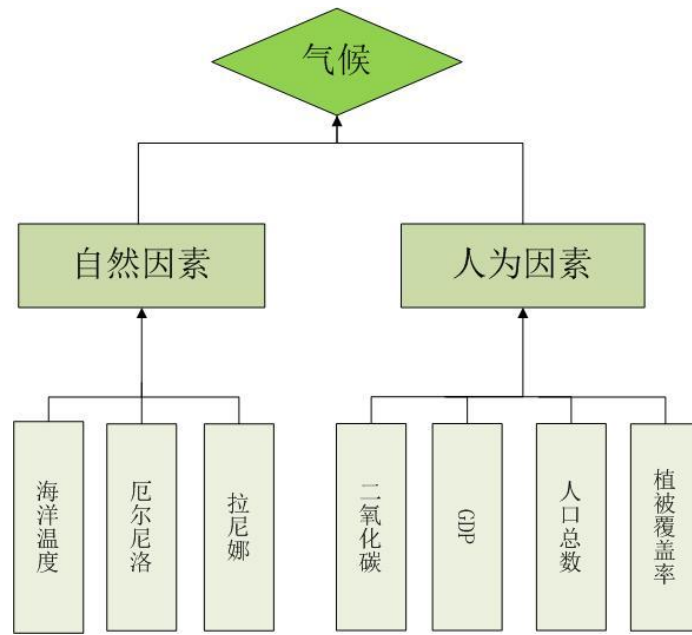


图 27 简化后的相关因子机理图

4.2.2 复杂气候模型—子模型的建立

基于 STF 为目标的多因素路径分析结果简化机理分析图，以权重相对较高的部分一级指标作为研究对象建立复杂气候模型—子模型的物理方程如下：

4.2.2.1 海洋温度模型

“海气耦合”理论认为在地球接收到来自太空的温度一定并且不考虑地表的吸热与散热的情况下，海洋热量与大气热量守恒。基于“非线性海-气耦合随机模式及绝热条件下的福克-普郎克方程”，海洋温度与大气温度之间的关系并不是简单两者正负反馈的问题，而是在海气内部各种因素以及随机因子的相互作用下产生的结果。

微分方程模型通常是运用所谓的平衡原理，即物资在某段时间内的变化量与其在这段时间内增加和减少的差处于平衡状态，因此利用平衡状态来建立海洋温度与大气温度之间的数学模型是十分合适的。

在大气中，根据热量平衡方程^[18]，若略去气温在经纬度上的水平运输量，则“地气系统”整体气温变化可以表示为：

$$\frac{\delta T}{\delta t} = \frac{H_G + H_L + H_R}{PC_p H} \quad (16)$$

式中： T 为气温； H_G 为感热通量； H_L 为潜热通量； H_R 为辐射差额； P 为空气密度； C_p 为定压比热。

对于海洋温度，若略去海水盐分，平液和消耗的影响，整体海洋温度的变化亦类似地可以写为：

$$\frac{\delta T_s}{\delta t} = \frac{H_G^w + H_L^w + H_R^w}{P^w C_p^w H} \quad (17)$$

式中： W 为含 W 上标的变量含义对于与大气温度中的表达式

在时间微段 dt 内，假设地球上“地气系统”与海洋系统吸放热平衡，不考虑时间微段中地球与太空发生热量交换，则在时间微段 dt 内两者体系的吸放热平衡，有如下平衡方程：

$$\frac{\delta T}{\delta t} + \frac{\delta T_s}{\delta t} = 0 \quad (18)$$

基于上述平衡方程，可以得到“地气系统”温度与海洋温度之间的微分关系如下：

$$\frac{\delta T}{\delta t} = -\frac{\delta T_s}{\delta t} \quad (19)$$

基于上述“地气系统”温度与海洋温度之间的微分关系得到如海洋温度数学模型如下：

$$V[(u, v), t] = -\iint \frac{\delta T_s}{\delta t} dudv \quad (20)$$

4.2.2.2 二氧化碳含量模型

二氧化碳对于地球的影响是基于地球系统的能量平衡：地球所有的能量来源均来自于太阳释放的短波辐射，到达地球后，一部分被地表（或大气层）反射（或散射）到宇宙中区，一部分被地表吸收。通俗来讲，就是地球晒了太阳后就会发生内部系统复杂的温度变化。然后地表升温之后就会对外放出长波辐射，放出的长波辐射一部分就散溢到太空中，一部分会被大气层中的温室气体加热，大气层就会被加热^[20]，加热后的大气层亦会重新散发出长波辐射对“地气系统”进行加热从而形成了“温室效应”现象。

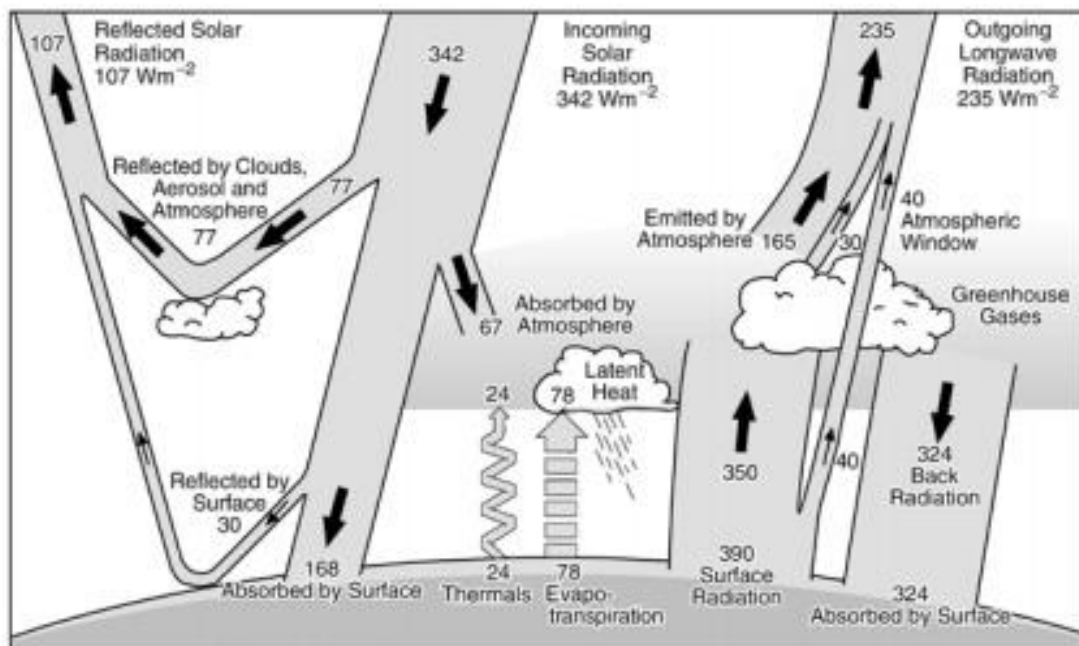


图 28 生态系统二氧化碳循环示意图

大气中的碳循环仿真系统模型:

$$\begin{aligned} M_{AT}(t) &= E(t) + \phi_{11}M_{AT}(t-1) + \phi_{21}M_{UP}(t-1) \\ M_{UP}(t) &= \phi_{12}M_{AT}(t-1) + \phi_{22}M_{UP}(t-1) + \phi_{32}M_{LO}(t-1) \\ M_{LO}(t) &= \phi_{23}M_{UP}(t-1) + \phi_{33}M_{LO}(t-1) \end{aligned} \quad (21)$$

大气中的二氧化碳主要来源于如下两类:

$$\begin{aligned} &\sum_{u=1}^{u=longitude} \sum_{v=1}^{v=latitude} Ec[(u,v),t] = \sum_{u=1}^{u=longitude} \sum_{v=1}^{v=latitude} E_{industry}[(u,v),t] \\ &+ \sum_{u=1}^{u=longitude} \sum_{v=1}^{v=latitude} E_{agriculture}[(u,v),t] \\ &\sum_{u=1}^{u=longitude} \sum_{v=1}^{v=latitude} Ec[(u,v),t] \text{—大气中二氧化碳总含量数学模型} \\ &\sum_{u=1}^{u=longitude} \sum_{v=1}^{v=latitude} E_{industry}[(u,v),t] \text{—工业部分二氧化碳总含量数学模型} \\ &\sum_{u=1}^{u=longitude} \sum_{v=1}^{v=latitude} E_{agriculture}[(u,v),t] \text{—农业产生二氧化碳总含量数学模型} \end{aligned} \quad (22)$$

温室气体总辐射排强迫变化量的表达式如下:

$$F(t) = k \left[\log_2 \left(\frac{\sum_{u=1}^{u=longitude} \sum_{v=1}^{v=latitude} E_c[(u,v),t]}{M_{AT}(1980)} \right) \right] + F_{EX}(t) \quad (23)$$

4.2.2.3 人口及经济发展模型

人口对于温度的影响主要是通过能源的消耗排入大量温室气体, 温室气体对于温度的影响机理如前文, 人口增长率随时间的变化数学模型如下:

$$g_L(t) = \frac{g_L(t-1)}{1 + \delta_L} \quad (24)$$

人口通过消耗能源与化石燃料来影响环境温度的变化, 基于上述人口增长率模型建立人口与大气温度之间的模型:

$$Z[(u,v),t] = Ng_L(t) \quad (25)$$

式中: N 为人口温度之间的函数。

4.2.2.4 厄尔尼诺以及拉尼娜周期变化模型

厄尔尼诺现象厄尔尼诺海流, 是太平洋赤道带范围内海洋和大气相互作用后失去平衡而产生的一种现象。正常情况下, 热带太平洋区域的季风洋流是从美洲走向亚洲, 使太平洋表面保持温暖, 给印尼周围带来热带降雨。

拉尼娜是指赤道太平洋东部和中部海面温度持续异常偏冷的现象 (与厄尔尼诺现象正

好相反)，是热带海洋和大气共同的产物。

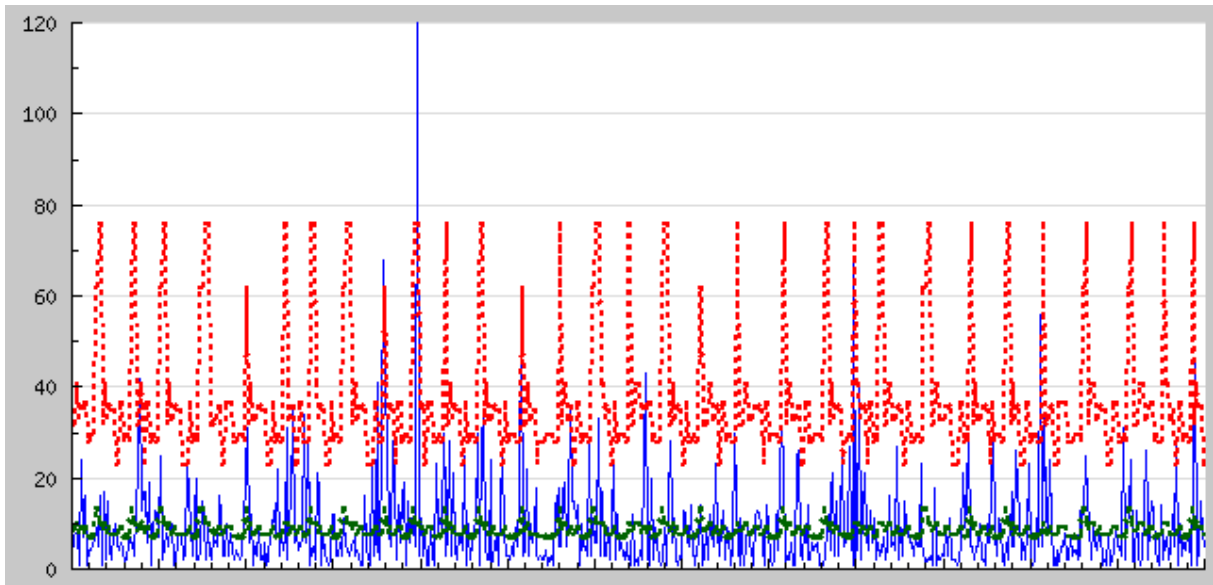


图 29 厄尔尼诺周期振荡图

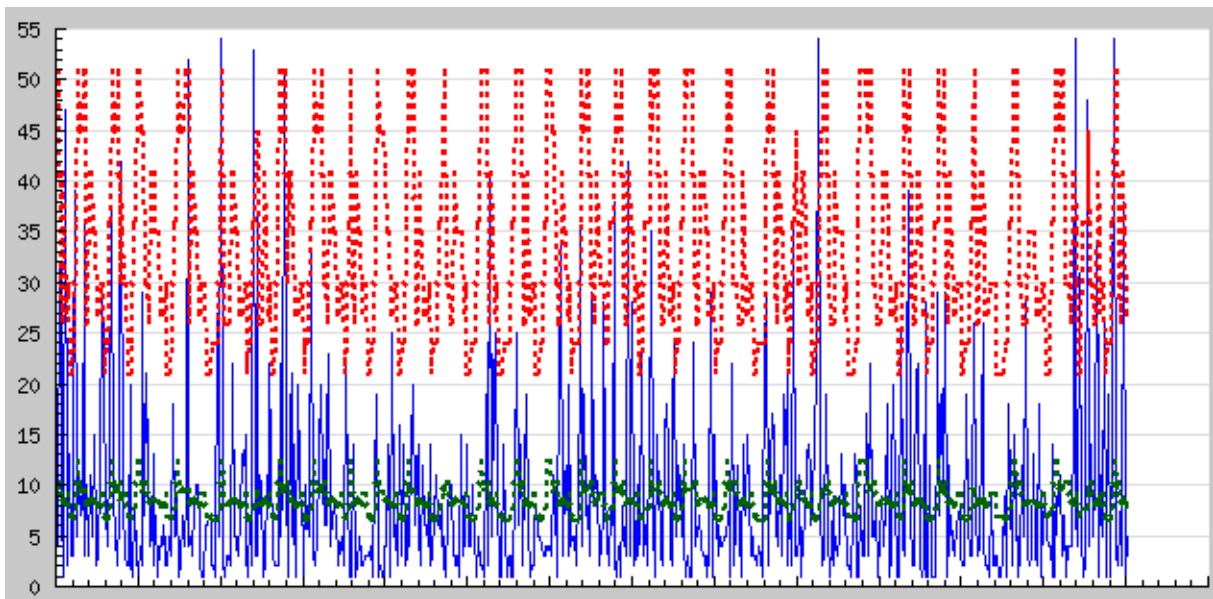


图 30 拉尼娜周期振荡图

由以上振荡周期图，厄尔尼诺现象和拉尼娜现象是具有明显的周期性现象，基于周期性原理可以对厄尔尼诺现象建模分析过程如下：

$$\begin{aligned}
 M^E[(u, v), t] &= \alpha \sin\left[\text{sen}\left(\frac{t-t_e}{T^E}\right) + \frac{\pi}{2}\right] \\
 M^L[(u, v), t] &= \beta \cos\left[\text{sfn}\left(\frac{t-t_l}{T^L}\right) + \frac{\pi}{2}\right]
 \end{aligned}
 \tag{26}$$

sen (θ) 函数说明—当 θ 为整数的时候函数取1, 当 θ 为小数的时候函数取0

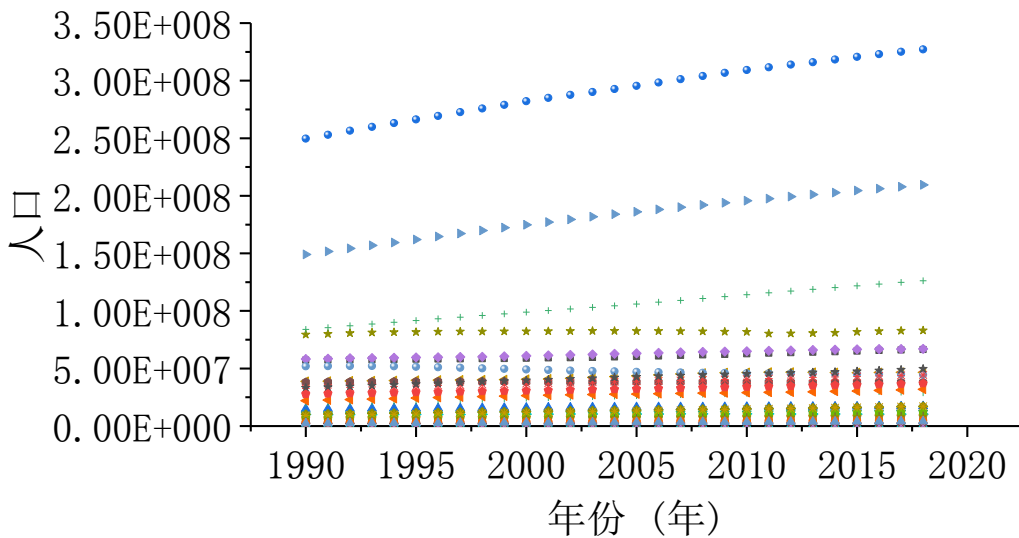
sfn (θ) 函数说明—当 θ 为整数的时候函数取0, 当 θ 为小数的时候函数取1

4.2.3 复杂气候模型—子模型数据的收集与预处理

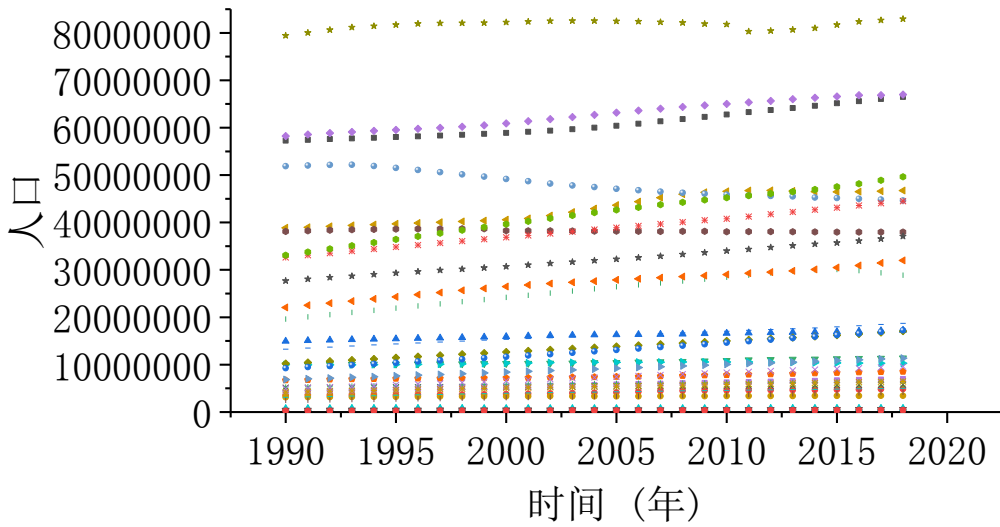
前面已经确定了影响全球气候变化的主要因素有人口数量、GDP、CO2 排放量和森林覆盖率, 本次工作收集到了亚洲、欧洲、非洲、北美洲、南美洲、大洋洲共 6 个洲在 1990 年—2016 年间 4 个主要因素的数据, 在每个洲中根据不同国家的贫富差距、面积大小, 地理位置等因素, 随机抽取出几个国家的四个因素的数据值, 6 个洲共抽取了 34 个国家, 并对其进行分析。

4.2.3.1 人口因素

在统计的 34 个国家中, 分析其在 26 年间的人口变化情况, 如图; 为更加清晰的观测到所有国家的人口变化速率, 对人口在 1.5×10^8 以下的国家另外进行统计, 如图。在此基础上统计出 34 个国家总的人口数量在 26 年间的变化情况, 如图。



(a)



(b)

图 31 34 个国家在 26 年间的人口变化情况 (颜色不同的点代表不同国家)

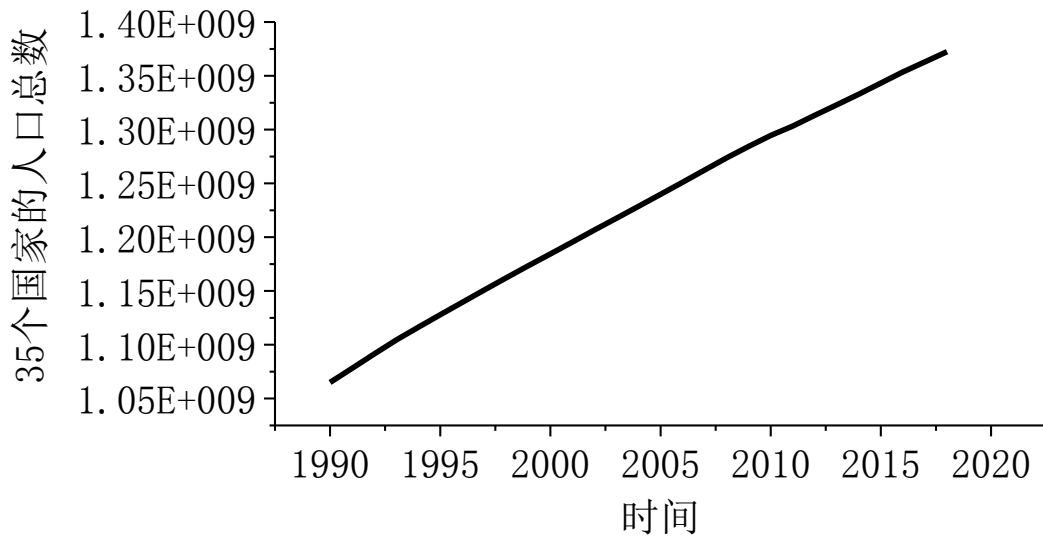
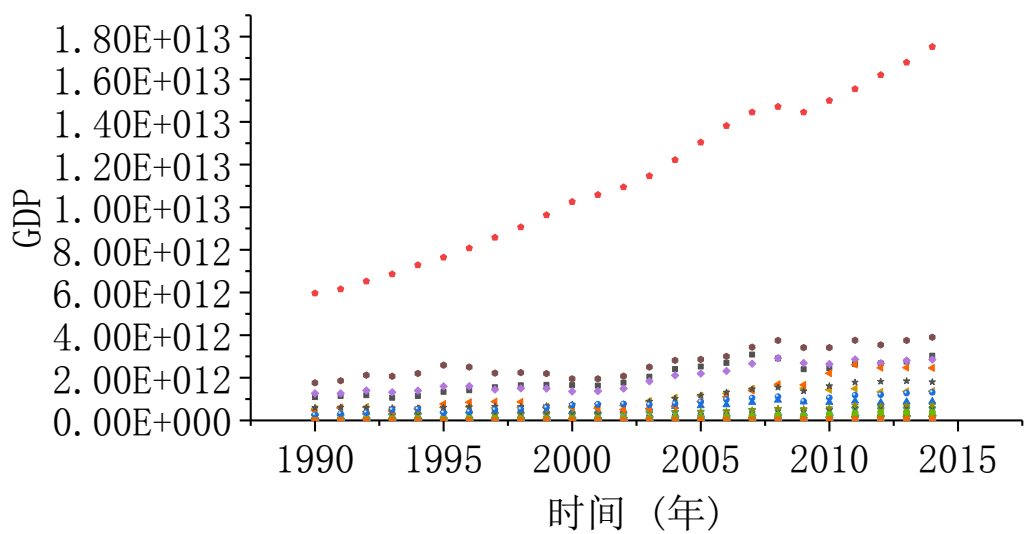


图 32 34 个国家在 26 年间的总人口变化情况

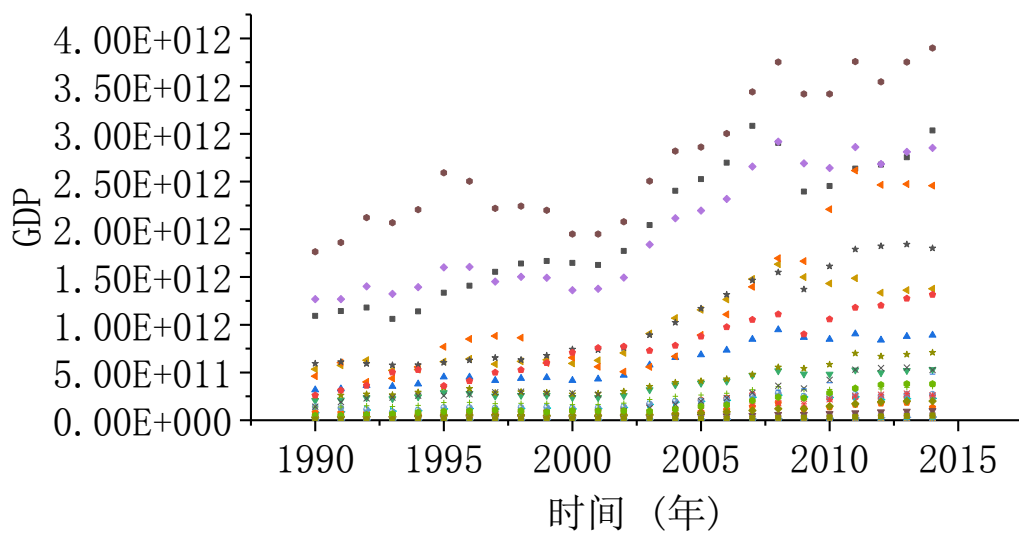
由图可知, 26 年间全世界范围内人口数量总体呈现上升趋势, 但该上升的趋势是较为平缓的。其中, 美国的人口增长速率是最快的, 在 34 个国家中只有两个国家的人口数量出现了负增长。总体来说, 人口基数大的国家人口的增长速率越快。

4.2.3.2 GDP

统计到 34 个国家在 26 年间国内生产总值的变化情况, 如图, 为更加清晰的观测到所有国家的 GDP 变化速率, 再作出除去 GDP 数值最大曲线的 GDP 波动情况图, 如图。除此之外, 再分析 34 个国家总的国内生产总值, 如图。



(a)



(b)

图 33 34 个国家在 26 年间的 GDP 值变化情况

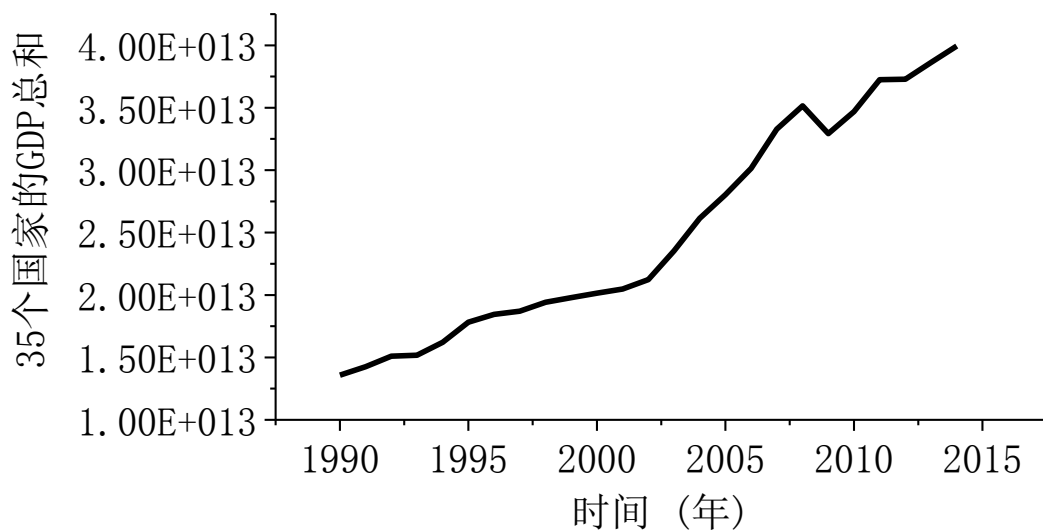


图 34 34 个国家在 26 年间总的 GDP 值变化情况

由图可知，34 个国家 GDP 数值是不不断波动的，在 1995 年—2002 年间局部区域的 GDP 数值有所降低，但总体来说呈现上升趋势的。其中，美国的 GDP 基数最大且增长速率最快，在 2008 年由于金融危机的原因，GDP 数值有所下降。34 个国家总的 GDP 值在 26 年间总体是不断上升的，在 2004 年前增长速率交缓，在 2004 年之后增长速度变快。

4.2.3.3 CO₂

统计出 34 个国家在 26 年间 CO₂ 排放量变化情况，如图。除此之外，再分析 34 个国家总的 CO₂ 排放量变化情况，如图。

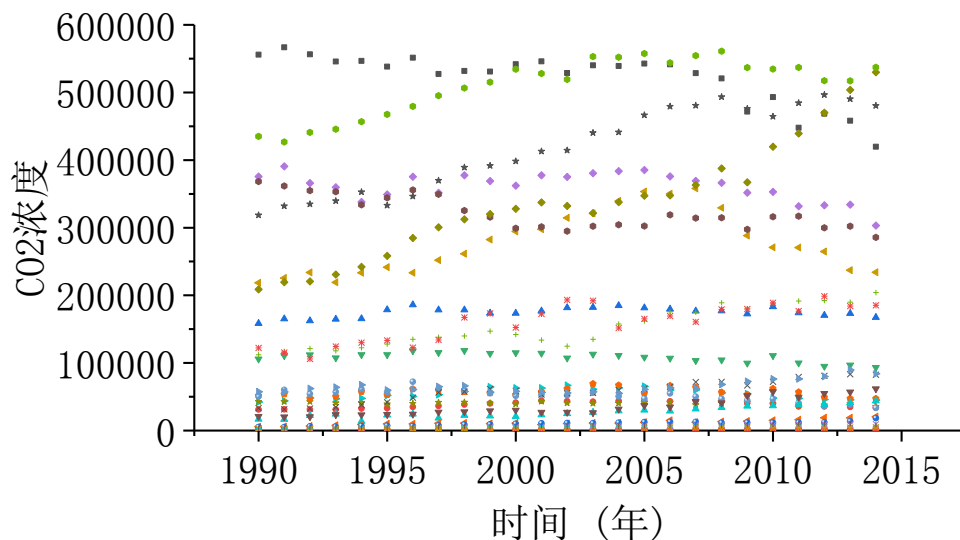


图 35 34 个国家在 26 年间的 CO₂ 排放量变化情况

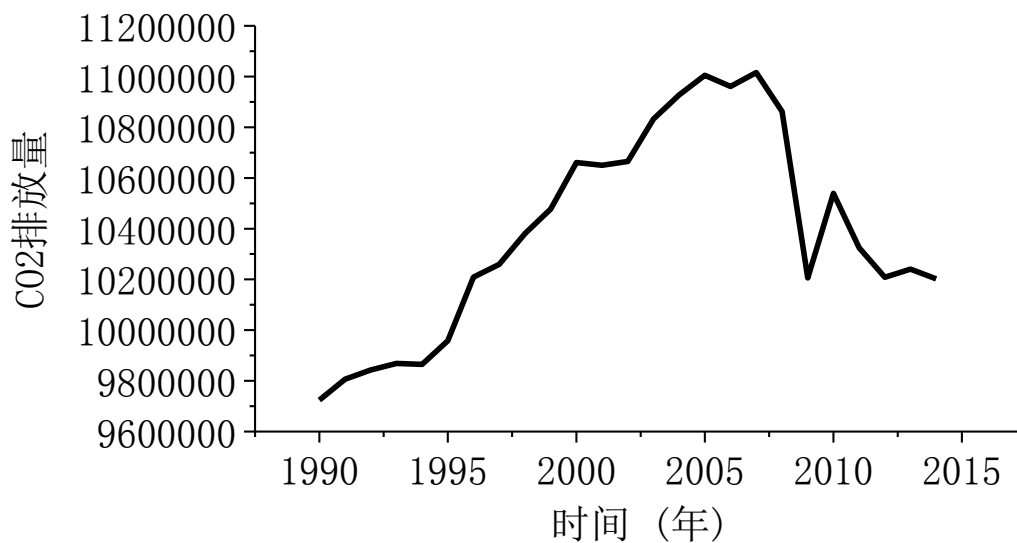


图 36 34 个国家在 26 年间总的 CO₂ 排放量变化情况

由图可知，CO₂ 排放量在 2007 年之前是不断上升的，在 2007 年之后，CO₂ 排放量急剧减小，随后出现了两次波动。

4.2.3.4 森林覆盖率

统计出 34 个国家在 26 年间森林覆盖率变化情况，如图。

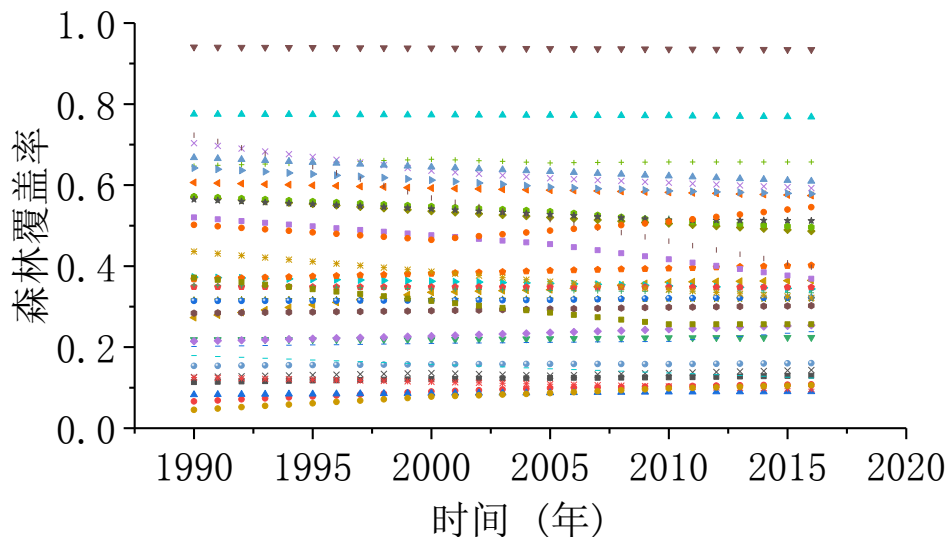


图 37 34 个国家在 26 年间的森林覆盖率变化情况

由图可知，各个国家的森林覆盖率在 26 年间的变化范围不大，且未出现波动。

4.2.4 基于能量守恒的多因素复杂气候模型的建立

基于地球生态系统整体能量守恒方程建立 $\nabla_Q[(u,v),t]$ 气候预测模型如下：

$$\nabla_Q[(u,v),t] = (1-\Phi)Q^+[(u,v),t] - Q^-[(u,v),t]$$

$Q^+[(u,v),t]$ -地球整体吸热时空函数

$Q^-[(u,v),t]$ -地球整体放热时空函数

Φ -地球吸热折折减系数，该系数与太阳的周期性活动有关

$$Q^+[(u,v),t] = \lambda_1 Z[(u,v),t] + \lambda_2 X[(u,v),t] + \lambda_3 C[(u,v),t] + M^E[(u,v),t]$$

$$Q^-[(u,v),t] = \eta_1 X[(u,v),t] + \eta_2 V[(u,v),t] + \eta_3 C[(u,v),t] + M^L[(u,v),t]$$

$$M^E[(u,v),t] = \alpha \sin\left[\text{sen}\left(\frac{t-t_e}{T^E}\right) + \frac{\pi}{2}\right]$$

$$M^L[(u,v),t] = \beta \cos\left[\text{sfn}\left(\frac{t-t_l}{T^L}\right) + \frac{\pi}{2}\right]$$

$\text{sen}(\theta)$ 函数说明—当 θ 为整数的时候函数取1，当 θ 为小数的时候函数取0

$\text{sfn}(\theta)$ 函数说明—当 θ 为整数的时候函数取0，当 θ 为小数的时候函数取1

式中： Z 为人口影响函数； X 为温室气体影响函数； C 为深林覆盖率影响函数； V 为海洋温度影响函数； M^E 为厄尔尼诺影响周期函数； M^L 为拉尼娜影响周期函数； Q^+ 为整体吸热函数； Q^- 为整体放热函数； U 为经度序列变量； V 为纬度序列变量； T 为时间序列变量； λ 为各吸热分项权重系数； η 为各放热分项权重系数； ξ 为整体吸热折减系数；

4.2.5 基于多元高次多项式非线性回归的模型简化与求解

基于子模型收集的数据利用机器学习中的多元高次多项式非线性回归的方法对基础数据进行学习，得到相关因素与温度之间的回归方式，基本形式如式。

$$Y = \beta_0 + \beta_1 X_1^k + \beta_2 X_2^k + \dots + \beta_p X_p^k + e \quad (27)$$

式中： $X_1^k X_2^k \dots X_p^k$ ； $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 为回归系数； e 为不可观测的随机误差； K 为多项式高阶次数。

(1) 划分基础数据集为训练集和测试集。为了测评机器学习所得数学模型的拟合程度，将数据库中的数据按照 3:1 的比例随机划分为“训练集”和“测试集”，“训练集”用于机器学习而“测试集”用于测评“训练集”所得数学模型的拟合程度，训练集并不参与机器学习的过程。

(2) 相关系数 R 。相关系数是用于反映变量场和结构响应之间相关关系密切程度的统计指标，各变量的相关系数与数学模型中各变量值系数呈正比。相关系数在 0-0.3 之间为弱相关，0.3-0.6 之间为弱相关，0.6-1.0 之间为强相关，相关系数表达式如式。

$$R = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (28)$$

式中： R 为相关系数； $\text{cov}(X, Y)$ 为 X 和 Y 的协方差； σ_X ， σ_Y 为标准差。

(3) 训练回归模型。基于 (1) 中划分的训练集，利用多元高阶多项式非线性回归方法对训练集进行机器学习，系数计算表达式如式，回归系数的求解采用以残差平方和最小为目标的 LS 估计法。

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (29)$$

式中： X 为变量场纯量矩阵； Y 结构响应矩阵。

(4) 用 ROC 曲线和精度指标（判定系数）测试线性回归模型的精度。将“测试集”中的变量场水平值输入“训练集”得出预测数学模型便可以得到“测试集”的预测值，将预测值与“测试集”的实际值做对比并且绘制两者的 ROC 曲线，判定系数可作为定量判断机器学习精度的指标，其计算表达式见式。ROC 曲线可以较为直观的判断拟合结果的好坏，用判定系数的值来直观的判定机器学习结果的好坏。

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (30)$$

式中： y_i 为待拟合数值； \bar{y} 为均值； \hat{y} 为拟合值。

4.2.6 模型的求解与结果

(1) 模型的求解

Step1: 气候的全面机理分析

对影响气候指标的因素建立三级嵌套指标，全面分析影响气候变化的原因，建立全面的机理分析结果。

Step2: 以 STF 为目标的多因素路径分析

利用带有约束条件的最小二乘回归算法进行各因素权重分析并且略去高阶微量，最后建立简化的机理分析结果。

Step3: 气候预测子模型的建立

建立高权重影响因素进行的复杂气候原理的子模型

Step4: 基于能量守恒原理的多因素复杂气候模型的建立

组合各类子模型建立多因素复杂气候模型如下：

$$\Delta_Q [(u, v), t] = (1 - \xi) Q^+ [(u, v), t] - Q^- [(u, v), t] \quad (31)$$

$$\Delta_Q = (1 - \xi) Q^+ - Q^- = \begin{cases} > 0, \text{升温} \\ = 0, \text{平衡} \\ < 0, \text{降温} \end{cases} \quad (32)$$

Step5: 组合模型的简化

将组合的各类非线性子模型进行简化，简化原理为多元高次多项式非线性回归模型如下：

$$Y = \beta_0 + \beta_1 X_1^k + \beta_2 X_2^k + \dots + \beta_p X_p^k + e \quad (33)$$

Step6: 模型简化前后对比

将组合复杂气候模型与简化后的模型求解并且预测未来 25 年的温度趋势，将两种模型进行对比分析，对比发现：简化后的模型虽然与复杂气候模型相比精度有所降低，但是其形式更加简单并且利于非专业人士对于气候变化趋势的理解。

问题二的解决过程流程图如下：

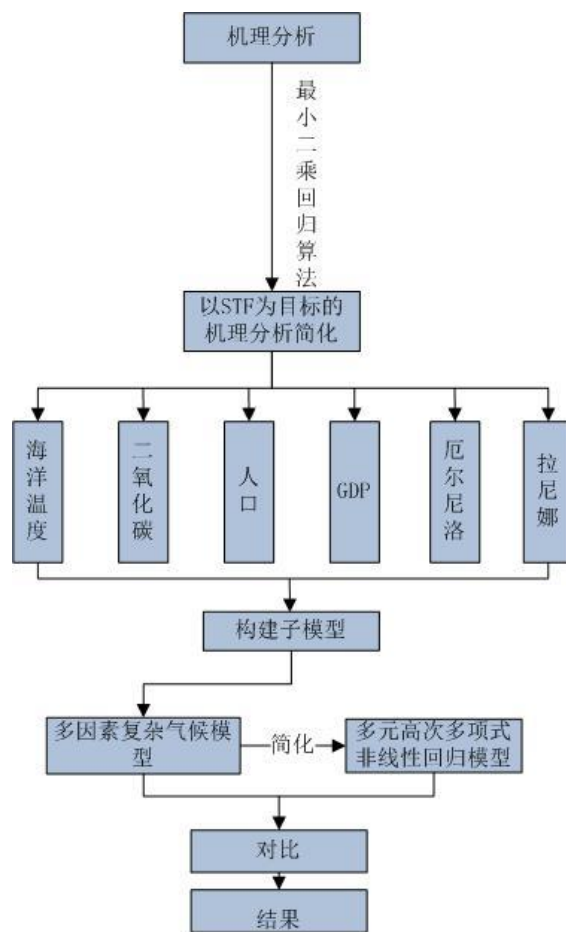


图 38 问题二解决过程流程图

(2) 结果

把上述的人口数量、GDP 和 CO₂ 排放量前一年到第二年的比值考虑为变化因子，分析 3 个因素在 26 年间的波动情况。

1

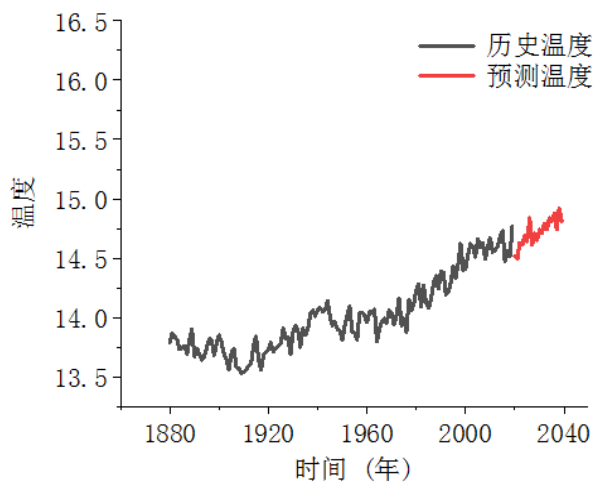


图 39 全球 25 年温度预测图

通过对 34 个国家 CO₂ 浓度、植被覆盖率、人口总数、GDP 等数据分析，应用基于能

量守恒的多元高次多项式气温预测模型得出在未来 25 年内的变化趋势如上图所示，全球气温呈现上升态势，在 1940 年左右出现突增，考虑到经济的发展影响到 GDP 的提高，从而影响到温室气体的排放。温室气体排放增长，下个世纪全球将出现缓慢但显著的变暖，由预测模型计算出在 2045 年左右全球平均气温将升高 0.6 摄氏度左右。

由于所考虑因素，以及不同的气候模型都对未来全球变暖的预测存在很大的差异。但全球气温仍呈现上升的趋势。根据资料查询，由于人为排放加速、气溶胶负荷减少和海洋环流条件的变化，1.5℃边界可能在 2030 年左右的一半时间内通过，2℃边界可能在 2045 年左右通过。同时变暖预测显示出一种“厚尾”分布，其变暖的可能性更大，远远超过气候模型预测的平均变暖量，其概率高于典型统计假设下的预期。

最近的评估表明，到 2100 年，全球气温将上升 4 摄氏度多一点。如果经济增长继续下去，并且没有实施有效的干预措施，就会超过工业化前的水平，实际气温上升可能会大或小几度，这取决于未来排放量的实际趋势以及对温室气体水平升高的气候反应仍然知之甚少。由此我们的预测全球温度在 2045 年上升 0.6℃是一个合理的预测结果。

4.3 问题三：考虑极端气候现象的“小波分析-越险指标”天气预测模型

4.3.1 极端天气机理分析

极端天气事件是一种在特定地区和时间(一年内)的罕见事件。“罕见”的定义有多种，但极端天气气象站气事件的罕见程度一般相当于观测到的概率密度函数小于第 10 个或第 90 个百分位点。按照定义，在绝对意义上，极端天气特征因地区不同而异。单一的极端事件不能简单地直接归因于人为气候变化，因为总是有一个有限的几率:极端事件可能会自然发生。当一种型态的极端天气持续一定的时间，如某个季节，它可归类于一个极端气候事件，特别是如果该事件产生一个平均极值或总极值(如:某个季节的干旱或暴雨)。



图 40 极端天气影响影响因子机理分析图

表 1 近年来极端天气总结

2019	北半球，美国中西部寒气逼人。1 月 30 日，美国芝加哥早上的气温降到-23℃，打破了 1966 年创纪录的-15℃。美国国家气象局数据显示，芝加哥风寒指数达到-52，即考虑风速影响，芝加哥最低温度相当于无风状态下的-52℃。明尼苏达州创造全美寒冷新“巅峰”，气温低至-38℃，风速 35 英里/小时，风寒指数-70，突破美国最低纪录。
	南半球，澳大利亚酷暑难耐。1 月 24 日，阿德莱德气温达到 46.6℃，打破 80 年来的纪录。怀阿拉 48.5℃、利溪 46.9℃、奥古斯塔港 49.1℃，全部打破历史最高纪录。高温造成大量野生动物死亡，其中澳大利亚眼镜狐蝠两天内死亡三分之一。
2018	7 月 15 日，日本全国的 927 个气象站中，有 200 个记录下 35 摄氏度以上的高温。
	7 月 3 日到 10 日期间，摩洛哥在 7 月 3 日经历了 43.4 摄氏度的创纪录高温。

	7月5日，阿尔及利亚沙漠地区的一个气象站录得 51.3 摄氏度的最高气温，创下该国历史纪录。
	美国西海岸受到炎热天气影响。7月8日，加州的一个气象站记录到 52 摄氏度的极端高温。洛杉矶地区创下一连串新的高温纪录，平均夜间气温达 26.1 摄氏度，郊区最高气温 48.9 摄氏度。
	西伯利亚北部地区的异常高温从六月一直延续到了七月。7月9日至16日期间，该地区连续五天经历了超过 30 度的高温，导致森林火灾风险加剧，电力供应受到影响。位于西伯利亚的克拉斯诺亚尔斯克的气温比往年平均高出 7 摄氏度，山火影响了超过 8 万公顷的森林。
	斯堪的纳维亚半岛遭遇持续高温，北极圈气温达到 30 摄氏度，挪威和芬兰分别报告了 33.5 和 33.4 摄氏度的创纪录高温，瑞典则在七月中旬发生了大约 50 起森林火灾。6月1日到7月16日之间的降雨量仅有 47 毫米。
2017	法国巴黎在热浪之后于 10 日遭遇有降水记录以来最大的一次暴雨，1 小时内降雨量达到 49 毫米，超过正常 7 月 3 个星期的降雨总量。
	6 月中下旬以来，印度多地进入季风雨季，雷电暴雨天气频发。印度媒体 10 日报道说，比哈尔邦和北方邦发生多起雷击事件，造成至少 19 人死亡。在遭受雷电袭击较严重的比哈尔邦，至少 16 人死亡，另有两人受伤严重。
2016	中国气温为历史第三高，而全国平均降水量为历史最多，达到 730 毫米，较常年偏多 16%。温度高、降水多、极端天气事件频发。去年 11 月至目前，中国平均气温为 1961 年来最高。
	全球共出现 93 次热带气旋，比正常情况高出 13%，其中包括导致海地约 1000 人遇难的飓风“马修”。
	全世界冰川平均萎缩约 1 米，为连续第 37 年萎缩。

4.3.2 小波分析的极端气候现象周期趋势分析

小波分析 (wavelet analysis) 或小波转换 (wavelet transform) 是指用有限长或快速衰减的，称为母小波 (mother wavelet) 的振荡波形来表示信号。该波形被缩放和平移以匹配输入的信号。

4.3.2.1 确定小波变换参数

用 ARIMA 模型对非平稳时间序列 $\{x_t\}$ ($t=1,2,\dots,N$) 建立预测模型之前，需要选择对 $\{x_t\}$ 进行分解的小波基函数和分解层数。选择正确的小波基函数对小波分解至关重要，由于不同的小波基在紧支性，正交性，对称性和平滑性上会描述出不同性质，因此不同的小波基函数对同一序列的分析会呈现出各种不同的效果。从应用预测的角度，结合各小波基函数的不同特征，通过对序列处理的理论误差分析可以判定是否选择了正确的小波基函数。

表 2 常用小波函数的性质

小波函数	harr	daubeclnes	symlets	meyer
正交性	有	有	有	有
紧支撑性	有	有	有	有

支撑长度	1	2n-1	2n-1	有限长度
对称性	对称	近似对称	近似对称	对称

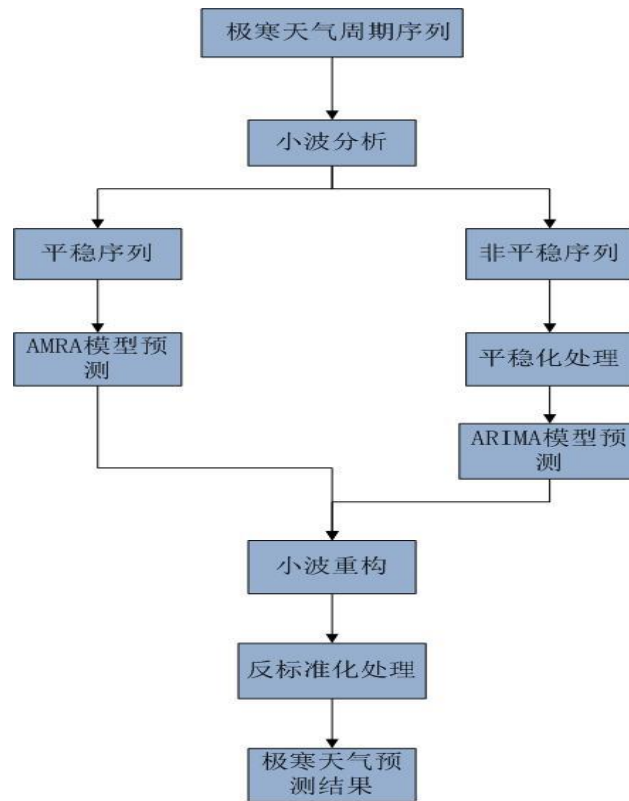


图 41 小波分析预测流程图

4.3.2.2 极寒天气周期预测函数

基于上述小波分析原理，对极端气候发生的周期进行预测，经过简单的一元非线性拟合得到极寒天气出现周期预测函数如下：

$$T_{cold} = k \ln^t + b \quad (34)$$

式中：k 为变化速率系数；b 为截距系数。

4.3.2.3 开关函数

基于上述小波分析原理得出的极寒天气预测周期模型可以得到预测极寒天气是否发生的“开关函数”如下：

$$\rho_{switch} = \sin\left[\text{sen}\left(\frac{t - t_{extreme}}{k \ln^t + b}\right) + \frac{\pi}{2}\right] \quad (35)$$

sen (θ) 函数说明—当θ为整数的时候，函数取1，当θ为小数的时候，函数取0

式中：t_{extreme} 为极端气候发生的起点时间值

4.3.3 基于越险风险指标的极寒天气阈值计算模型

利用前述小波分析的极端气候现象周期趋势分析可以对极端天气出现的年份进行预测，但是并未考虑到极端天气气候时空序列相关上下限阈值，由极端天气的定义可知极端天气上下限定限阈值与时空序列函数的最低温度和平均温度的差值波动有关。

某一时空特征下极寒天气越险严重程度数学模型如下：

$$\nabla T_{mean-min}[(u, v), t] = T_{mean}[(u, v), t] - T_{min}[(u, v), t] \quad (36)$$

式中： $T_{mean}[(u, v), t]$ 为时空序列温度平均值； $T_{min}[(u, v), t]$ 为时空序列低温值。

提出严重度指标 S_K 表示发生极寒天气时的温度低于平均温度的严重程度，计算数学模型如下：

$$S_k = \sum_{u=1}^{longitude} \sum_{v=1}^{latitude} F_L[(u, v), t] \nabla T_{mean-min}[(u, v), t] \quad (37)$$

式中： F_L 为时空权重系数。

4.3.4 “小波分析-越险指标”的极端气候现象天气预测修正模型

通过对极寒天气发生的年份进行小波分析，然后通过非线性拟合得到极端天气发生时间周期函数，然后得到极寒天气是否发生的开关函数。通过对时空序列函数的最低温度和平均温度的差值波动进行越险指标分析得到极端天气发生严重指标，对上述开关函数和严重指标进行组合得到“小波分析-越险指标”的极端气候现象天气预测模型如下：

$$H = \rho_{swi} S_c \quad (38)$$

4.3.5 模型的求解与结果

(1) 基于本问题中的“小波分析-越险指标”的极端气候现象天气预测修正模型表达式可知，修正模型的开关函数直接与温度时空变化特征相关，因此极寒天气的出现频率与全球温度变化趋势有着直接的关联。

基于 Python 的编程语言对上述修正模型进行程序，输出极寒天气的程度的振荡曲线如下：

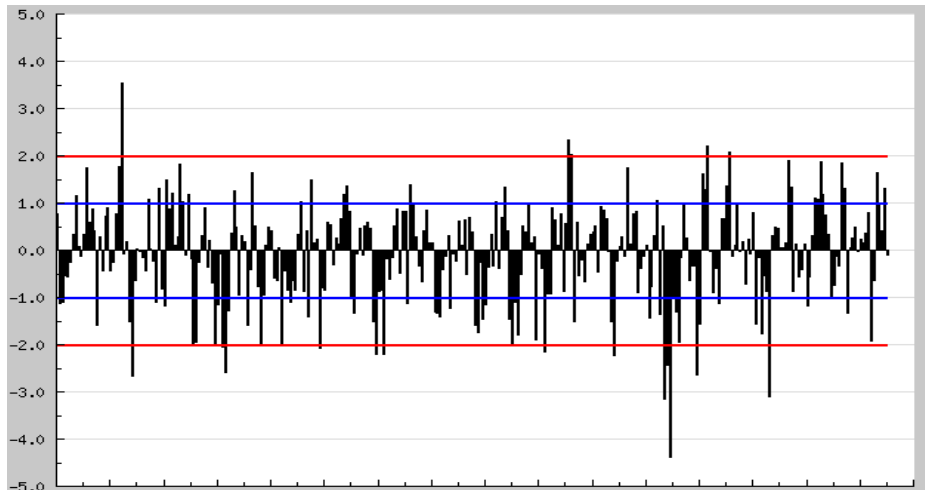


图 42 极寒天气振荡曲线

由于第二问已经得出全球气温整体变暖的趋势，因此基于上图可以得出随着全球气温变暖会导致极寒天气的程度加大。

基于 Python 的编程语言对上述修正模型进行程序，输出极寒天气出现的频数统计图如下：

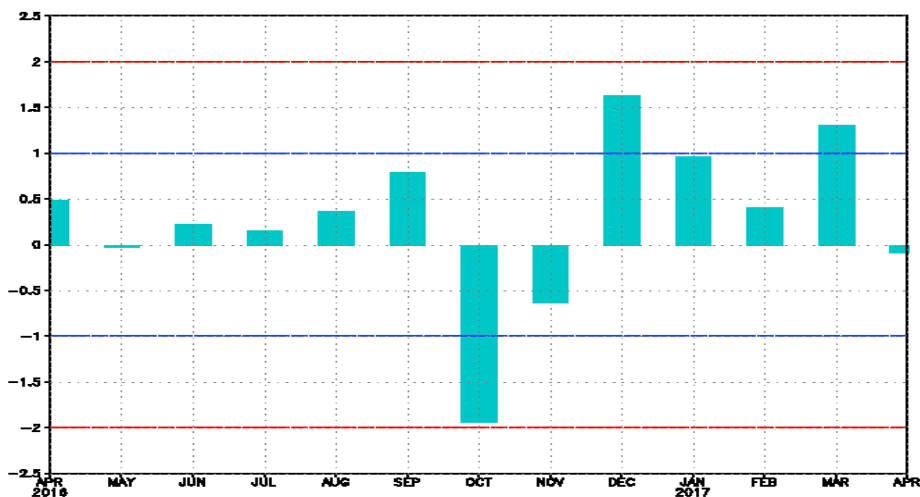


图 43 极寒天气频数统计图

由上图可知，随着全球气温变暖，全球极寒天气出现的频率增加。

(2) 全球变暖与局地极寒天气的出现并不矛盾，因为基于以上两图的分析，随着全球气温的升高，极寒天气出现的频率与剧烈程度反而升高，其原因在于气候是一个长时间内气候要素和天气现象的平均统计状态，而极寒天气是一种短时间内的天气现象，全球变暖与局部极寒的关系相当于“整体—局部”的关系，并不矛盾。

4.4 问题四：表征全球变暖的新概念“多因素气候决策指标”的提出

4.4.1 分析数据

对 1950 年—2019 年间的厄尔尼诺和拉尼娜事件的发生情况进行了统计，包括发生的起止时间、峰值时间、峰值强度和强度等级，如表所示。

表 3 ENSO 历史事件统计表

1950 年以来厄尔尼诺/拉尼娜事件

序号	起止年月	长度 mon	峰值时 间	峰值强 度 °C	强度等 级	事件类 型	
暖 事 件	1	1951.08— 1952.01	6	1951.11	0.8	弱	东部型
	2	1957.04— 1958.07	16	1958.01	1.7	中等	东部型
	3	1963.07— 1964.01	7	1963.11	1.1	弱	东部型
	4	1965.05— 1966.05	14	1965.11	1.7	中等	东部型
	5	1968.10— 1970.02	17	1969.02	1.1	弱	中部型
	6	1972.05— 1973.03	11	1972.11	2.1	强	东部型
	7	1976.09— 1977.02	6	1976.10	0.9	弱	东部型
	8	1977.09— 1978.02	6	1978.01	0.9	弱	中部型
	9	1979.09— 1980.01	5	1980.01	0.6	弱	东部型
	10	1982.04— 1983.06	15	1983.01	2.7	超强	东部型
	11	1986.08— 1988.02	19	1987.08	1.9	中等	东部型
	12	1991.05— 1992.06	14	1992.01	1.9	中等	东部型
	13	1994.09— 1995.03	7	1994.12	1.3	中等	中部型
	14	1997.04— 1998.04	13	1997.11	2.7	超强	东部型
	15	2002.05— 2003.03	11	2002.11	1.6	中等	中部型
	16	2004.07— 2005.01	7	2004.09	0.8	弱	中部型
	17	2006.08— 2007.01	6	2006.11	1.1	弱	东部型
	18	2009.06— 2010.04	11	2009.12	1.7	中等	中部型

序号	起止年月	长度 mon	峰值时 间	峰值强 度 °C	强度等 级	事件类 型	
19	2014.10—2016.04	19	2015.12	2.8	超强	东部型	
20	2018.09—2019.06	10	2018.11	1.0	弱	中部型	
冷 事 件	1	1950.01— 1951.02	12	1950.01	-1.4	中等	东部型
	2	1954.07— 1956.04	22	1955.10	-1.7	中等	东部型
	3	1964.05— 1965.01	9	1964.11	-1.0	弱	东部型
	4	1970.07— 1972.01	19	1971.01	-1.6	中等	东部型
	5	1973.06— 1974.06	13	1973.12	-1.8	中等	中部型
	6	1975.04— 1976.04	13	1975.12	-1.5	中等	中部型
	7	1984.10— 1985.06	9	1985.01	-1.2	弱	东部型
	8	1988.05— 1989.05	13	1988.12	-2.1	强	东部型
	9	1995.09— 1996.03	7	1995.11	-0.9	弱	东部型
	10	1998.07— 2000.06	24	2000.01	-1.6	中等	东部型
	11	2000.10— 2001.02	5	2000.12	-0.8	弱	中部型
	12	2007.08— 2008.05	10	2008.01	-1.7	中等	东部型
	13	2010.06— 2011.05	12	2010.12	-1.6	中等	东部型
	14	2011.08—2012.03	8	2011.12	-1.1	弱	中部型
	15	2017.10—2018.03	6	2018.01	-0.8	弱	东部型

4.4.2 统计分析及结论

(1) 分析

把表中的强度等级划分为4个指标：弱、中等、强、超强；分别对应数值1、2、3、4。绘出1950年~2018年间厄尔尼诺和拉尼娜事件的强度等级数值图。

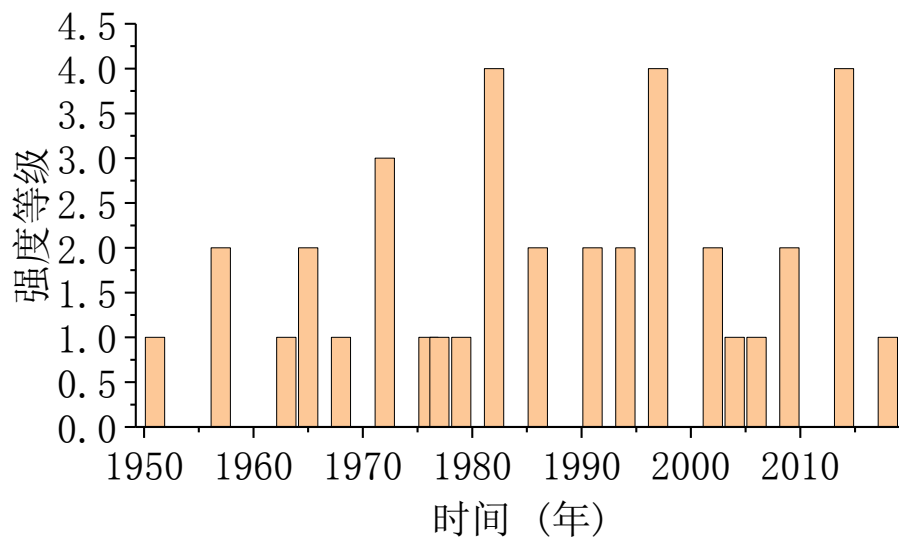


图 44 厄尔尼诺天气频数统计图

由图可知，厄尔尼诺事件在 1975 年之前，发生次数较少且强度普遍偏弱。在 1975 年之后，共发生 3 次超强的厄尔尼诺事件，且发生的次数有所增加。

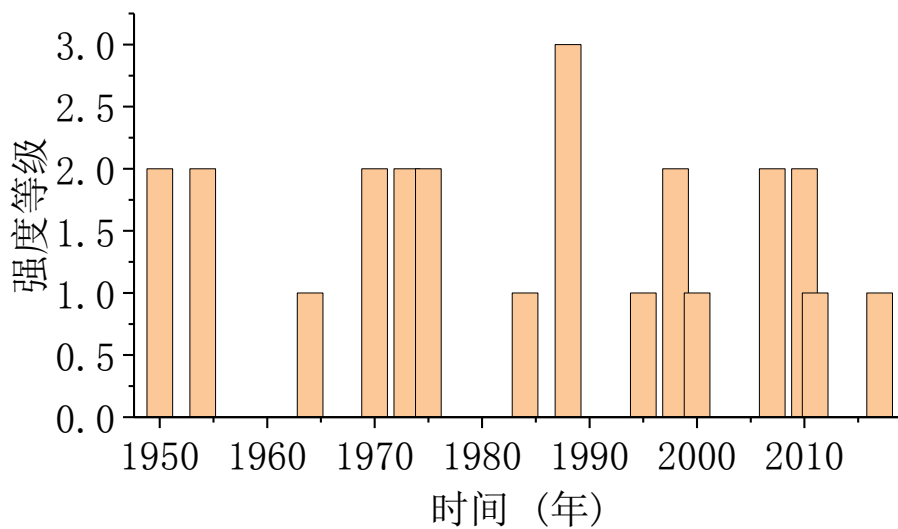


图 45 拉尼娜天气频数统计图

由图可知，拉尼娜事件在 1975 年之前，发生次数较少。在 1975 年之后，发生 1 次超强的拉尼娜事件，且发生的次数明显增加。

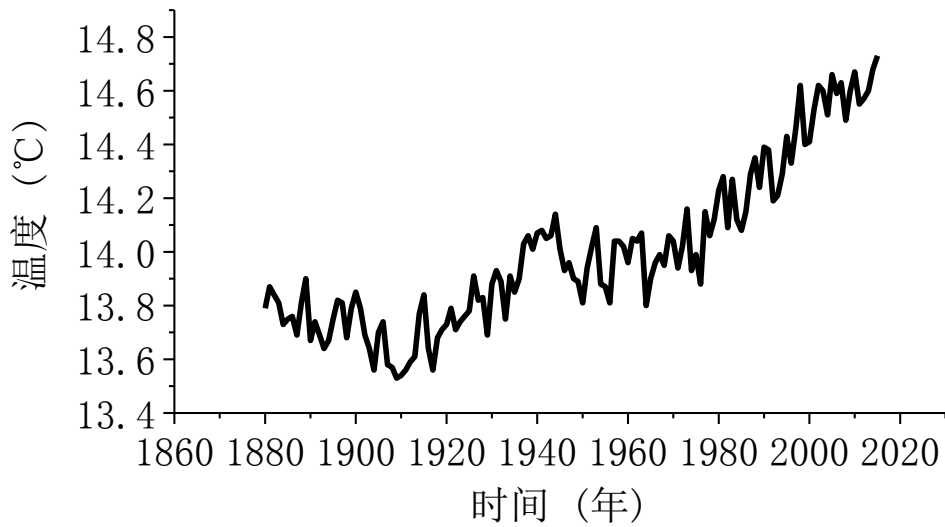


图 46 全球年平均温度变化趋势

由上图可以，全球年平均气温趋势虽然在局部上有波折，但是全球温度的整体态势是升温的。

(2) 结论

从气候角度研究全球温度变化的时候需要全球范围长时间的数据的观测积累，在长时间的温度变化中，会出现波动的现象，在某些时间段的温度可能会降低，但是根据前文的研究结果中，整体是会出现上升趋势的。综上所述，全球变暖与“某地今年特别冷”之间并没有直接相关性。

由于影响温度的因素众多，如经纬度、海拔高度、CO₂ 浓度、人口数量、GDP、森林覆盖率等众多因素，在统计这些关键因素的时候，由于横跨的时间段长，全球“站点”数目极多，导致需要收集的数据量极大。在对全球气候变化情况作出预测的时候，需要收集众多因素在过去相当长一段时间的数据，但过去在这方面的时空数据并不完整，给统计计算带来了极大的困难。同时，这些影响因素之间也会相互影响，同时生态系统也是一个循环的过程，导致了预测气候变化是一个及其复杂的过程，综上所述，以“多因素气候决策指标”作为反映气候变化趋势和复杂性的新概念。

五 模型的评价

5.1 模型的优点

(1) LSTM 神经网络预测模型可以解决传统神经网络模型无法处理长时间序列的难题。

(2) 基于多元高次多项式非线性回归的简化模型相较于传统模型相比更加容易被大众理解。

(3) “小波分析-越险指标”天气预测修正模型的组成模型充分显示了极寒天气发生的概率和剧烈程度内在机理。

5.2 模型的缺点

基于多元高次多项式非线性回归的简化模型精准度高度依赖于回归分析的数据集，当数据集来源不准确的时候回归模型准确性就会大大降低。

六 参考文献

- [1]齐庆华,蔡榕硕,郭海峡.中国东部气温极端特性及其气候特征[J].地理科学,1340-1350.2019,
- [2]李牧原,罗德海.冬季北极增暖与中纬度环流和极寒天气之间的联系:经向位涡梯度的关键作用[J].中国科学:地球科学,1335-1345,2019,
- [3]陆晴,闫冰,赵东升.1961-2016年江西省气候时空变化特征[J].水土保持研究,166-173.2019,
- [4]刘政阳,李挺宇.全球气候变暖趋势急剧加速[J].生态经济,1-4.2019,
- [5]Junming Fan,Hui Hong,Hongguang Jin. Life cycle global warming impact of CO₂ capture by in-situ gasification chemical looping combustion using ilmenite oxygen carriers[J]. Journal of Cleaner Production,2019,234.
- [6]Paolo Fastelli,Monia Renzi. Exposure of key marine species to sunscreens: Changing ecotoxicity as a possible indirect effect of global warming[J]. Marine Pollution Bulletin,2019.
- [7]Xiao-Yong Wang,Han-Seung Lee. Effect of global warming on the proportional design of low CO₂ slag-blended concrete[J]. Construction and Building Materials,2019,225.
- [8]Buddhi Wijesiri,An Liu,Nian Hong,Panfeng Zhu,Bo Yang,Xu Zhao,Ashantha Goonetilleke. Rethinking hydrocarbons build-up on urban roads: A perspective on volatilisation under global warming scenarios[J]. Environmental Pollution,2019,252.
- [9]Dongdong Yang,Hua Zhang,Jiangnan Li. Changes in concentrations of fine and coarse particles under the CO₂-induced global warming[J]. Atmospheric Research,2019,230.
- [10]Corentin Iltis,Philippe Lou pre,Karolina Pecharov  Denis Thi ry,S bastien Zito,Benjamin Bois,J r me Moreau. Are life-history traits equally affected by global warming? A case study combining a multi-trait approach with fine-grain climate modeling[J]. Journal of Insect Physiology,2019,117.
- [11]Alvaro Z niga-Reinoso,Reinhard Predel. Past climatic changes and their effects on the phylogenetic pattern of the Gondwanan relict *Maindronia* (Insecta: Zygentoma) in the Chilean Atacama Desert[J]. Global and Planetary Change,2019,182.
- [12]Jean P. Palutikof,Sarah L. Boulter,Frank Stadler,Ana C. Perez Vidaurre. Tracking the progress of climate change adaptation: An Australian case study[J]. Environmental Science and Policy,2019,101.
- [13]金浩宇,鞠琴,曲珍,董小涛,郝振纯.基于集成方法的长江源区未来气候变化预测研究[J/OL].水力发电:1-7[2019-09-22]
- [14]程婷婷,马姗姗,张思晴,王猛.1960—2018年北京市延庆区气候变化特征分析[J/OL].现代农业科技,142-143.2019
- [15]王丽娜,祝青林,王岩,周百慧.2018年大连地区极端天气气候事件及其影响[J/OL].现代农业科技,156-157.2019
- [16]陈志伟,康建成,顾成林,徐勇,汤明,陆凯.近33a西北太平洋极端海表温度事件的变化特征及与ENSO的关系[J].海洋环境科学,221-232.2019,
- [17]杨朝虹.西北太平洋热带气旋对流与降水特征研究[D].兰州大学,2018.
- [18]田宝强.长江中上游极端强降水频次变异规律[A].中国气象学会.第32届中国气

象学会年会 S4 东亚气候变异成因和预测[C].中国气象学会:中国气象学会,2015

[19]薛红喜,孟丹,吴东丽,王连喜,李琪.1959~2009 年宁夏极端温度阈值变化及其与 AO 指数相关分析[J].地理科学, 380-385.2012,

[20]林霄沛,许丽晓,李建平,罗德海,刘海龙.全球变暖“停滞”现象辨识与机理研究[J].地球科学进展, 995-1000.2016,

七 附录

(1) 部分 LSTM 机器学习代码

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import torch
from torch import nn
from torch.autograd import Variable

#LSTM (长短期记忆网络)
data_csv = pd.read_csv('C:/Users/DZF/Desktop/LSTM/data.csv', usecols=[1])
#pandas.read_csv读取时空序列数据

plt.plot(data_csv)
plt.show()
#数据预处理
data_csv = data_csv.dropna()
dataset = data_csv.values
dataset = dataset.astype('float32')
max_value = np.max(dataset)
min_value = np.min(dataset)
scalar = max_value-min_value
dataset = list(map(lambda x: x/scalar, dataset)) #将数据标准化到0~1之间

def create_dataset(dataset, look_back=2):#look_back 以前的时间步数用作输入变量来预测下一个时间段
    dataX, dataY= [], []
    for i in range(len(dataset) - look_back):
        a = dataset[i:(i+look_back)] #i和i+1赋值
        dataX.append(a)
        dataY.append(dataset[i+look_back]) #i+2赋值
    return np.array(dataX), np.array(dataY) #np.array构建数组

data_X, data_Y = create_dataset(dataset)
#data_X: 2*142    data_Y: 1*142

#划分训练集和测试集，70%作为训练集
train_size = int(len(data_X) * 0.7)
test_size = len(data_X)-train_size

train_X = data_X[:train_size]
train_Y = data_Y[:train_size]

test_X = data_X[train_size:]
test_Y = data_Y[train_size:]

train_X = train_X.reshape(-1, 1, 2) #reshape中，-1使元素变为一行，然后输出为1列，每列2个子元素
train_Y = train_Y.reshape(-1, 1, 1) #输出为1列，每列1个子元素
test_X = test_X.reshape(-1, 1, 2)
```

2 M-K 非参数统计检验

```
import numpy as np
def s(inputdata):
    #输入numpy数组
    n=inputdata.shape[0]
    t=0
    for i in np.arange(n):
        if i <=(n - 1):
            for j in np.arange(i+1,n):
                if inputdata[j]> inputdata[i]:
                    t=t+1
                elif inputdata[j]< inputdata[i]:
                    t=t-1
                else:
                    t=t
    return t

data=rd(r'D_S.tif')
#data为S，这里是针对一个单波段tif图像
n=10
var=n(n-1)(2n+5)/18
#n为时间序列的长度；var为方差
sv=np.sqrt(var)
#sv为标准差
r,c=data.shape
z=np.zeros(data.shape)
#由于时间久远，只用了for循环处理较慢，可用numpy矩阵运算效率更高
for i in range(r):
    for j in range(c):
        if data[i][j]>0:
            z[i][j]=(data[i][j]-1)/sv
        elif data[i][j]<0:
            z[i][j]=(data[i][j]+1)/sv

def beta(inputdata):
    n=inputdata.shape[0]
    t=[]
    for i in np.arange(n):
        if i <=(n - 1):
            for j in np.arange(i+1,n):
                t.append((inputdata[j]-inputdata[i])/((j-i)*1.0))
    return np.median(t)
```

```

import numpy as np
def Hurst(x):
    #x为numpy数组
    n=x.shape[0]
    t=np.zeros(n-1)    #t为时间序列的差分
    for i in range(n-1):
        t[i]=x[i+1]-x[i]
    mt=np.zeros(n-1)    #mt为均值序列,i为索引,i+1表示序列从1开始
    for i in range(n-1):
        mt[i]=np.sum(t[0:i+1])/(i+1)

    #Step3累积离差和极差,r为极差
    r=[]
    for i in np.arange(1,n):    #i为tao
        cha=[]
        for j in np.arange(1,i+1):
            if i==1:
                cha.append(t[j-1]-mt[i-1])
            if i>1:
                if j==1:
                    cha.append(t[j-1]-mt[i-1])
                if j>1:
                    cha.append(cha[j-2]+t[j-1]-mt[i-1])
        r.append(np.max(cha)-np.min(cha))
    s=[]
    for i in np.arange(1,n):
        ss=[]
        for j in np.arange(1,i+1):
            ss.append((t[j-1]-mt[i-1])**2)
        s.append(np.sqrt(np.sum(ss)/i))
    r=np.array(r)
    s=np.array(s)
    xdata=np.log(np.arange(2,n))
    ydata=np.log(r[1:]/s[1:])

    h,b= np.polyfit(xdata,ydata, 1)
    return h

```