

基于ARIMA的多元线性回归模型的全球气候变暖分析

周雨蕙, 唐旗英

上海理工大学, 健康科学与工程学院, 上海

收稿日期: 2023年1月12日; 录用日期: 2023年3月13日; 发布日期: 2023年3月21日

摘要

全球气候变暖已成为国际社会广泛关注的问题, 本文利用ARIMA和多元线性回归模型来分析全球的温度变化和影响气温的因素。首先我们对1992年到2022年全球平均气温的12个月份数据, 使用SPSS软件进行可视化与预处理, 对异常值的波动进行总结并建立数学模型。然后我们开发一种机器学习方法对火山爆发、森林火灾、新冠肺炎和太阳活动异常四种自然灾害的影响进行分析, 给出了长期线性趋势, 得出自然灾害对气候变暖的影响。接着对于影响全球气温变化的原因, 我们从降水量、海洋表面温度、全球平均温度、二氧化碳浓度、地球吸放热等因素进行考虑, 利用逐步回归法通过SPSS进行相关性检验求解多元线性微分方程, 得出地球放热对温度影响最大。最后我们对模型进行了优化和推广, 将ARIMA模型与多元线性回归预测模型进行结合, 利用极寒天气情况对模型进行推广, 检验模型的灵敏度, 最终的拟合效果较好, 应用场景较广。

关键词

全球变暖, 机器学习方法, ARIMA, 多元线性回归模型

Global Warming Analysis Based on ARIMA Multiple Linear Regression Model

Yuhui Zhou, Qiyang Tang

School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Jan. 12th, 2023; accepted: Mar. 13th, 2023; published: Mar. 21st, 2023

Abstract

Global warming has become a concern of the international community. In this paper, ARIMA and

multiple linear regression models are used to analyze the global temperature change and the factors that affect the temperature. Firstly, the 12-month data of global mean temperature from 1992 to 2022 were visualized and preprocessed by SPSS software, and the fluctuations of outliers were summarized and mathematical models were established. Then, we developed a machine learning method to analyze the impact of four natural disasters, volcanic eruption, forest fire, COVID-19 and abnormal solar activity, and gave a long-term linear trend to get the impact of natural disasters on climate warming. Then, for the reasons affecting the change of global temperature, we considered precipitation, ocean surface temperature, global average temperature, carbon dioxide concentration, earth heat absorption and release and other factors. We used the step regression method to solve the multivariate linear differential equation by SPSS correlation test, and concluded that the earth heat release had the greatest influence on temperature. Finally, we optimized and extended the model, combined the ARIMA model with the multiple linear regression prediction model, promoted the model with the extreme cold weather, and tested the sensitivity of the model. The final fitting effect was good and the application scenarios were wide.

Keywords

Global Warming, Machine Learning Methods, ARIMA, Multiple Linear Regression Model

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来, 世界各地相继出现了较为严重的温度变化。从工业时代开始, 人们对自然的无节制的能源利用以及人类活动的快速扩张使全球变暖问题日益严峻, 二氧化碳等温室气体浓度急速增长。这些温室气体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, 所以大气层内存在着大量累积热量, 使得地球表面气温度升高, 全球气候变暖。全球气候变暖可能引发一系列问题, 如干旱、风雹、暴雨、高温等极端天气, 不仅打破了稳定的生态系统, 也威胁着可持续发展与人类正常的生活。在这一领域中, 学者如刘政阳、李挺宇等[1]确切指出中国地表气温上升趋势明显。陈振坤、贾积身等[2]学者以加拿大地区为例利用非线性回归模型浅谈了其在气候变暖方面的应用。张宸豪等[3]基于大数据分析提出了气候变化模型。全球气候变暖已经成为国际社会广泛关注的问题。本文希望利用 ARIMA 和多元线性回归模型来分析全球的温度变化和影响气温的因素, 为改善全球变暖和实现绿色复苏的方向提供参考。

2. ARIMA 模型建立

2.1. 模型假设

- 1) 假设未来地球的生态系统能够保持相对稳定, 不会发生较大较严重的板块运动和地质活动;
- 2) 假设人类科技不会出现重大突破, 未来主要使用能源仍与目前现有的能源相似;
- 3) 假设短期不会出现新的影响地球气候的人为和非人为因素。

2.2. 数据预处理

我们选取了截止至 2013 年的 100 个城市的温度变化情况的数据。其中 81 个是北半球的城市, 19 个

为南半球的城市, 由于早期的数据比较多, 每月的平均温度也存在缺失。所以我们先进行了数据预处理工作, 使用序列平均值和邻近点的线性趋势方法, 将缺失值替换, 并剔除无效数据, 其中的平均温度不确定度表示统计和空间欠采样效应的 95% 置信区间。然后, 我们对这 100 个城市近 30 年的 3 月份上升温度进行了分析。并在网上收集了 1992 年到 2022 年全球平均气温数据, 并以十年作为一个节点进行对比处理, 并且我们关注到 3 月这个特殊的时间点温度, 可能和厄尔尼诺现象有关。

我们首先使用 matlab 提取出 1993、2003、2013 年这间隔 10 年的数据, 然后对他们的过去 30 年的 3 月份数据进行处理, 初步分析一下各地区 3 月份的一个温度变化情况, 为了更加直观的体现, 我们整理出相应结果并做成色阶图(图 1)。

City	The square of the increase in March over 30 years					
	Country	Latitude	Longitude	1993	2003	2013
Abidjan	Ivory Coast	5.63N	3.23W	0.487204	0.013689	0.011664
Addis Abe	Ethiopia	8.84N	38.11E	3.104644	1.545049	1.283689
Ahmadab.	India	23.31N	72.52E	8.678916	16.21673	19.9809
Aleppo	Syria	36.17N	37.79E	28.9444	14.0775	7.480225
Alexandria	Egypt	31.35N	30.16E	3.896676	1.803649	5.731236
Ankara	Turkey	39.38N	33.29E	37.36877	11.4921	8.357881
Baghdad	Iraq	32.95N	45.00E	18.5072	8.614225	7.546009
Bangalore	India	12.05N	77.26E	6.677056	2.719201	5.103081
Bangkok	Thailand	13.66N	99.91E	6.0025	0.632025	1.669264
Belo Horiz	Brazil	20.09S	44.36W	1.44	1.008016	0.426409
Berlin	Germany	52.24N	13.14E	16.62193	42.35406	0.497025
Bogotá	Colombia	4.02N	74.73W	0.068121	0.244036	0.150544
Bombay	India	18.48N	72.68E	0.5476	0.795664	1.679616
Brasília	Brazil	15.27S	47.50W	4.549689	0.423801	0.101761
Cairo	Egypt	29.74N	31.38E	10.3298	4.682896	14.1677
Calcutta	India	23.31N	88.25E	7.918596	9.659664	32.7184
Cali	Colombia	4.02N	76.34W	0.044944	0.114921	0.321489
Cape Tow	South Afri	32.95S	18.19E	0.331776	1.155625	0.133225
Casablanc	Morocco	32.95N	6.70W	8.928144	24.59168	8.219689
Changchu	China	44.20N	125.22E	82.64628	53.1441	84.16228
Chengdu	China	31.35N	103.66E	7.761796	6.906384	19.9809
Chicago	United Sta	42.59N	87.27W	10.16334	30.4704	3.356224
Chongqin	China	29.74N	107.08E	5.089536	4.397409	34.00056
Dakar	Senegal	15.27N	17.50W	0.003481	3.767481	0.097969
Dalian	China	39.38N	120.69E	26.84276	17.2142	34.2342

Figure 1. The display of March temperature data in some cities over the past 30 years

图 1. 部分城市过去 30 年的 3 月份气温数据的展示情况

然后, 我们采集了全球相对气温的数据, 并使用 excel 提取出 2013 年到 2022 年 12 个月份的数据情况, 再进行可视化处理, 做出了这十年期间 3 月份的全球相对气温变化情况, 并把其中 2014 年全球相对温度变化最明显的一年, 单独做出了 2014 年的表征。

从图 2 可以看出 2014 年的变化最为显著, 从图 3 可以看出 3 月份的变化趋势是最为陡峭的从 0.306 突变到 1.052。

接着, 对问题做进一步分析, 我们收集了全球气候周期的异常变化数据。对数据进行简单的预处理后, 导入 matlab 的 mat 中, 绘制了从 1992 年到 2022 年的每年对应月份平均温度值曲线, 还包括百分点为 2.5%、17%、83%、97.5% 时的年期间计算的年度周期曲线, 还从 NCDC 中提取了与上述年度周期有关的异常情况。

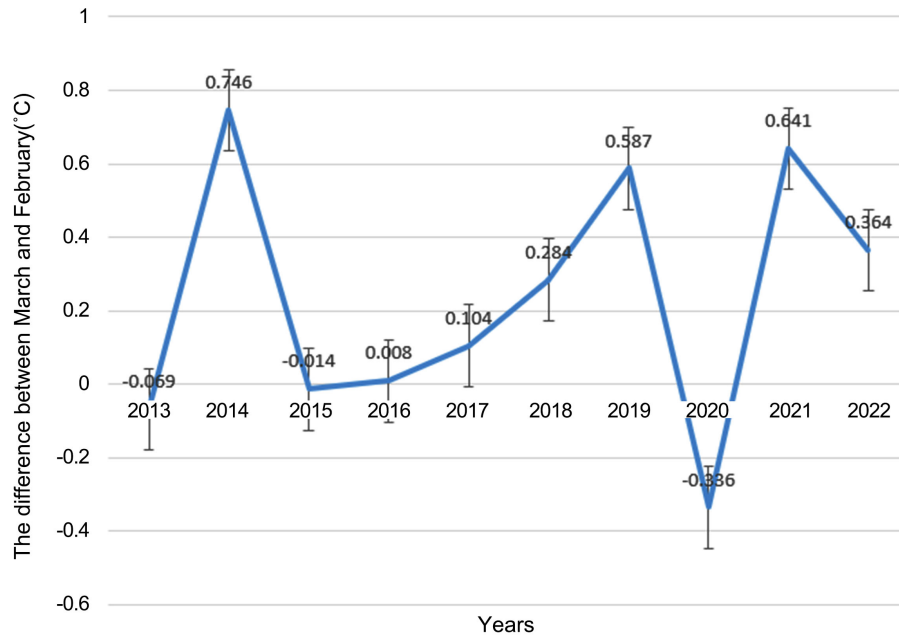


Figure 2. The March change in global relative temperature from 2013 to 2022
图 2. 2013 年到 2022 年, 全球相对气温在 3 月份的变化情况

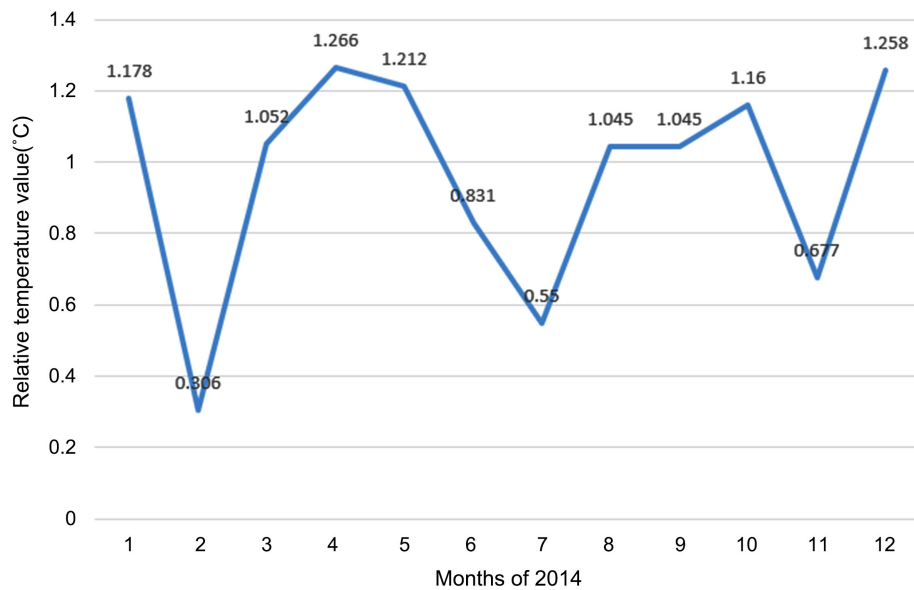


Figure 3. Analysis of global relative temperature in 2014
图 3. 2014 年的全球相对气温分析

通过可视化后, 我们可以观察到 3 月份的全球气温是上升变化最明显的(图 4)。

从图 5 可以看出整体波动的趋势[4]呈现逐年上升的趋势, 这也证明了我们的结论是正确的。

对这些数据进行处理分析得出 2022 年 3 月全球气温的上升的确是比以往任何 10 年期间观测到的升幅都要大, 原因可以从年度周期的异常情况可以看出, 异常值从 1880 年到 2022 年的波动趋势一直在上升。此外, 在我们研究的 1992 年到 2022 年, 这个年度周期中三月份的全球平均温度一直是异常的, 会明显高于其它月份。

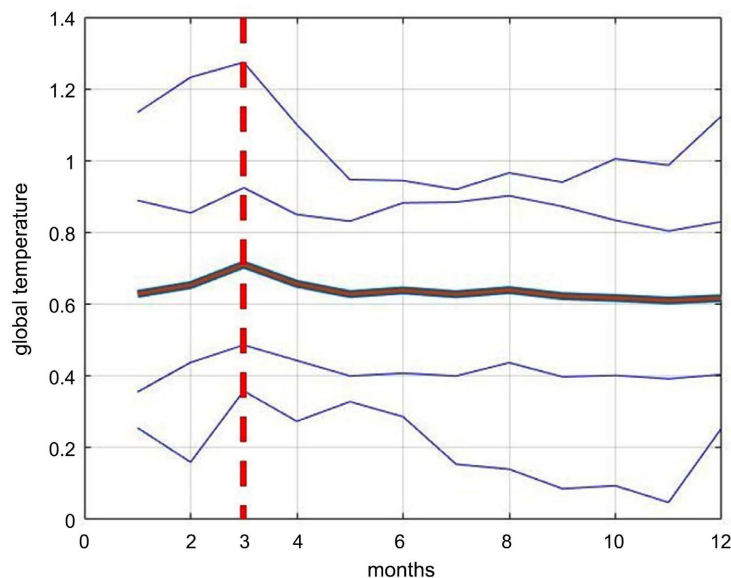


Figure 4. The global temperature curve is 97.5%, 83%, 17%, 2.5%, while the thickest dark red curve in the middle is the NCDC global Mean Temperature (1992:2022)

图 4. 97.5%、83%、17%、2.5%的全球气温曲线，而中间最粗的暗红曲线是 NCDC 全球平均气温(1992:2022)

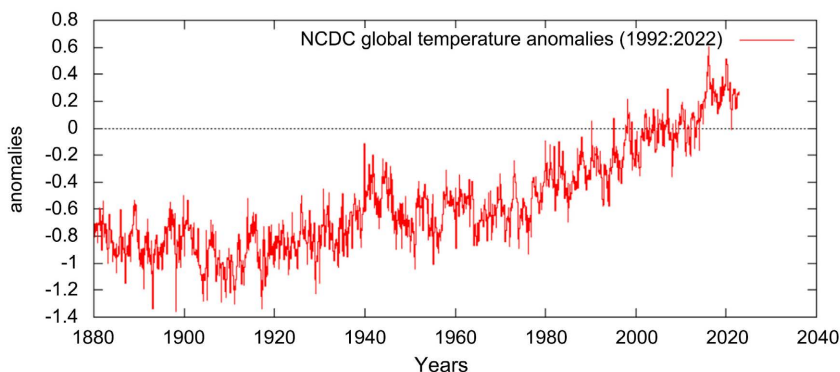


Figure 5. The red line represents the global temperature anomaly curve

图 5. 红色线波动表征的是全球气温异常情况曲线

2.3. 模型的建立

由于收集到的数据中有少数的缺失，对于缺失值的处理，由于不是大量数据缺失的情况，所以我们采用邻近点的线性趋势进行插值，可以得到更好的替换缺失值效果。

完成数据预处理后，通过 spss 软件绘制出时间序列图：

从图 6 可知，相对温度的整体波动趋势轻微上移，且在每年的三月左右，相对温度会出现一定范围内的极大值，该相对温度明显高于其他月份，因此数据表现出一定的季节性成分。

自回归模型只能适用于预测与自身历史数据相关性较强的数据，经过我们的检验发现数据符合这一标准，我们选取 ARIMA 模型，表达式如下

$$y_t = \delta + \sum_{i=1}^p \gamma_i \chi_{t-i} + \varepsilon \quad (1)$$

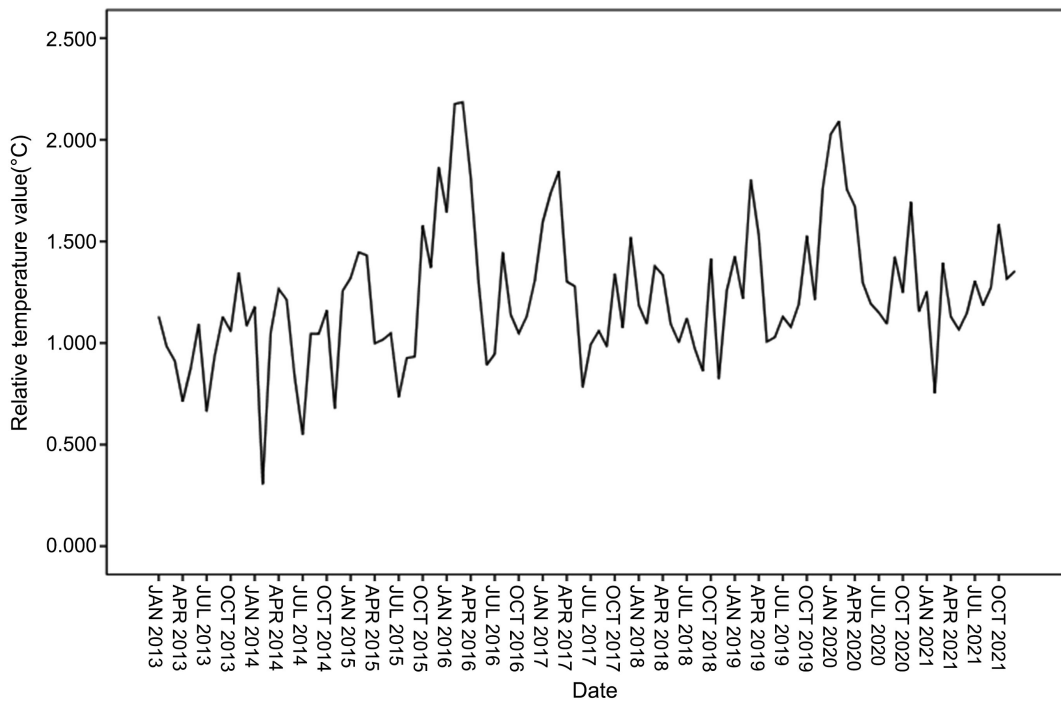


Figure 6. Time series graph of global mean relative temperature change (2013~2021)
图 6. 全球平均相对温度变化的时间序列图(2013~2021)

$$y_t = \delta + \sum_{i=1}^p \lambda_i y_{t-i} + \varepsilon_t \tag{2}$$

其中, δ 为常数, p 为阶数, ε 为误差, γ_i 为自相关系数。

分析过程与第一个模型类似。

从表 1 中, 我们可以分析出平稳的 R 方也是接近于 1 的, 同时正态化 BIC 也是比较小的, 这两个指标都体现了 ARIMA 模型的合理性。

Table 1. Model fit degree
表 1. 模型拟合度

拟合统计	平均值	标准误差	最小值	最大值	5%	10%	25%	50%	75%	90%
平稳 R 方	0.563	0	0.563	0.563	0.563	0.563	0.563	0.563	0.563	0.563
R 方	0.478	0	0.478	0.478	0.478	0.478	0.478	0.478	0.478	0.478
RMSE	0.249	0	0.249	0.249	0.249	0.249	0.249	0.249	0.249	0.249
MAPE	15.507	0	15.507	15.507	15.507	15.507	15.507	15.507	15.507	15.507
MaxAPE	76.570	0	76.570	76.570	76.570	76.570	76.570	76.570	76.570	76.570
MAE	0.186	0	0.186	0.186	0.186	0.186	0.186	0.186	0.186	0.186
MaxAE	0.671	0	0.671	0.671	0.671	0.671	0.671	0.671	0.671	0.671
正态化 BIC	-2.522	0	-2.522	-2.522	-2.522	-2.522	-2.522	-2.522	-2.522	-2.522

此外, 从表 2 中可以看出, 对残差进行 Q 检验得到的 p 值为 $0.655 > 0.05$, 即无法拒绝原假设, 认为残差是白噪声序列, 因此 ARIMA 模型也能很好的识别全球气温相对变化的数据。

Table 2. Model statistics
表 2. 模型统计

模型拟合度统计		杨-博克斯 Q (18)			
预测变量数	平稳 R 方	统计	DF	显著性	离群值数
0	0.563	13.240	16	0.655	3

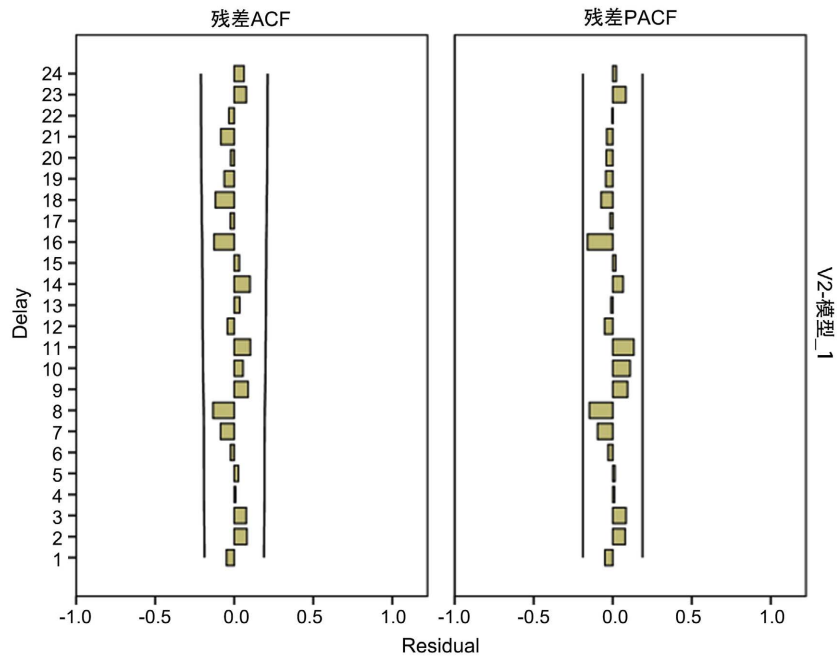


Figure 7. Residual ACF and PACF of data in the ARIMA model
图 7. ARIMA 模型中数据的残差 ACF 和 PACF

从残差的 ACF 和 PACF 图形中(图 7)可以看出, 所有滞后阶数的自相关系数和偏自相关系数均和 0 没有显著的差异。因此, ARIMA 模型更具有识别全球气温相对变化的数据的能力。

使用 spss 软件得出真实值、拟合值、预测值图(图 8)。

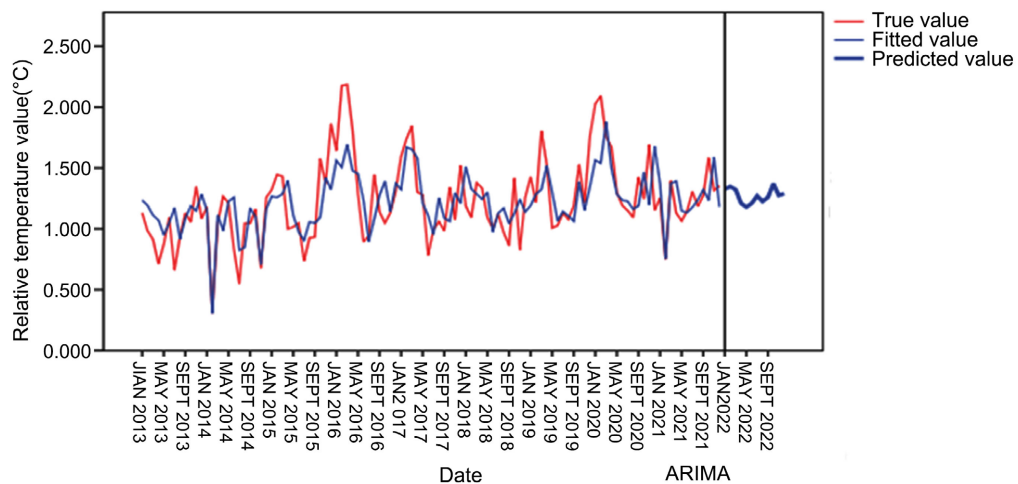


Figure 8. ARIMA model true value, fitting value, predicted value graph
图 8. ARIMA 模型真实值、拟合值、预测值曲线图

由于真实数据和拟合数据的时序图十分接近, ARIMA 模型的拟合效果更佳。在预测数据中持续了原始序列稳定变化, 这说明 ARIMA 模型更能对该全球相对气温数据进行预测。

2.4. 模型的可行性与合理性分析

我们使用 ARIMA 模型进行预测分析, 分析结果显示 R 方非常接近 1, 残差 ACF 与 PACF 只有极少数的数据点显著异于 0, 这说明该模型的使用是比较合理的。再使用 ARIMA 模型进行预测, 把 spss 中预测的结果导入到 excel 中, 接着使用 matlab 进行数据处理提取出 2016 年到 2250 年的预测结果, 从预测结果中发现全球平均温度在 2050 年达到了 16.6367°C, 而 2100 年将达到 17.5877°C, 这说明 2050 年与 2100 年达不到 20.00°C。ARIMA 预测模型中观测点的平均温度达到 20.00°C 时, 大概是在 2227 年, 该年的全球平均温度可以达到 20.0033°C (表 3)。

Table 3. Forecast results of the ARIMA model for 2016~2250

表 3. 2016 年~2250 年 ARIMA 模型的预测结果

年份	全球平均气温(°C)	年份	全球平均气温(°C)
2030	16.25633	2232	20.09842
2050	16.63667	2233	20.11742
2100	17.58767	2234	20.13642
.....		2235	20.15542
2225	19.96533	2236	20.17442
2226	19.98433	2237	20.19350
2227	20.00333	2238	20.21250
2228	20.02233	2239	20.23158
2229	20.04133	2240	20.25058
2230	20.06042	
2231	20.07942	2250	20.44067

通过我们利用指数平滑模型的简单季节性模型同时进行预测, 来比较模型的好坏。根据拟合度信息、平稳的 R 方、正态化 BIC (BIC 准则), 我们对两个模型进行一个初步的比较。由于正态化 BIC 的值, 越小的越好, 可以初步评估出 ARIMA 模型分析的更加准确。此外, 比较两个模型的残差 ACF 和 PACF 图形能够很直观的发现, ARIMA 模型的所有滞后阶数的自相关系数和偏自相关系数均和 0 没有显著的差异, 而指数平滑模型的简单季节性模型是有部分数据显著异于 0 的。

最后我们还可以分析拟合与预测的结果来评估两个模型的准确性, 从拟合结果图直观上分析, 可以判断出在早期时间范围内的拟合过程中, ARIMA 是拟合的更加准确的, 而在后期的时间序列范围内, 两者拟合结果准确度都不错。

因此, ARIMA 模型比简单季节性模型更加具有识别全球气温相对变化的数据的能力, 我们使用该模型预测是较为合理的。

3. 自然灾害对全球气温的影响

由于评估自然灾害等因素对全球气温的影响具有很强的现实意义, 本文将从新冠疫情, 森林火灾, 火山爆发, 太阳活动异常等可能的自然灾害对我们的多元非线性回归模型进行优化。

3.1. 新冠疫情

各国为了应对新冠病毒的全球流行颁布的一系列的措施, 世界人口流动性下降, 出行燃料的消耗量减少, 污染水平大幅下降, 间接性的降低了空气污染物和温室气体的释放, 对全球气候变化产生了一定影响。如新冠疫情爆发较为严重的 2020 年, 温室气体氮氧化物排放量下降了 30% 左右, 通过开发一种机器学习方法, 以得出全球空气质量监测站表面各气体浓度的变化。

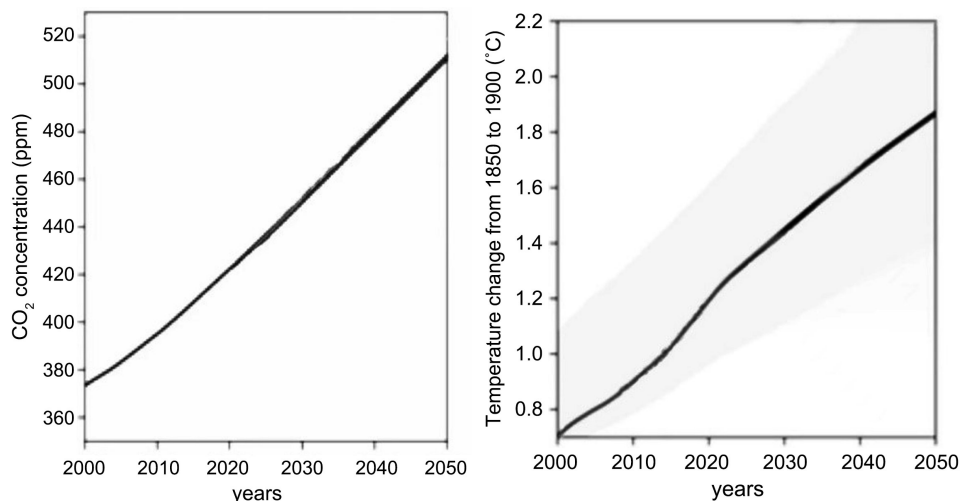


Figure 9. Estimated changes in carbon dioxide concentrations and climate response
图 9. 二氧化碳浓度和气候响应的估计变化

二氧化碳浓度和气候响应的估计变化如图显示, 如不对碳排放加以控制, 气候增长将与二氧化碳浓度呈正相关, 同时图中灰色部分显示的是蒙特卡罗的抽样不确定度(图 9)。

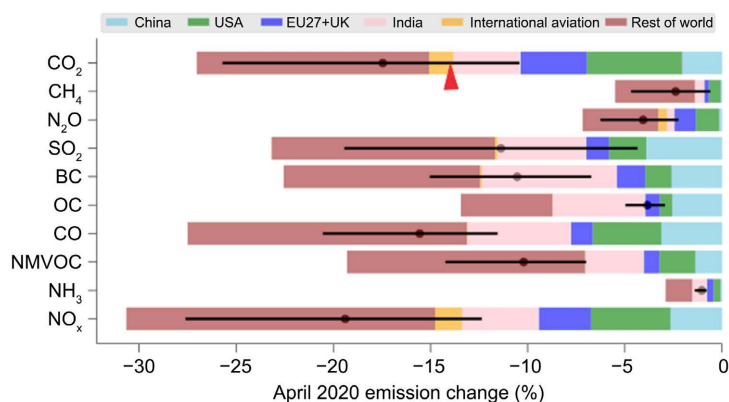


Figure 10. Global average emission reductions of different gases in April 2020 compared to the most recent year for China, the US, the EU and the UK, India, International aviation, and the rest of the world

图 10. 中国, 美国, 欧盟和英国, 印度, 国际航空, 其余国家 2020 年 4 月不同气体与最近一年相比的全球平均减排量

我们选取新冠疫情较为重要的时间节点 2020 年对主要气体排放进行归纳[5], 其中二氧化碳与基线相比, 二氧化碳的减少将带来更长时间的降温, 降温幅度约为 0.01 ± 0.005 摄氏度(图 10)。

新冠肺炎的传播会受到气候因素的影响。根据相关数据, 在 2020~2025 年期间, NO_x 下降高达 30%,

将带来高达 0.01 摄氏度的短期降温, 这大部分是由于对流层臭氧的减少。NO_x 也对硝酸盐气溶胶的减少产生了较小的变暖效应。气候保护政策措施很有可能将全球气温变化保持在高于工业化前 1.5 摄氏度的限制范围内, 到 2050 年节省约 0.3 摄氏度的未来变暖。

3.2. 太阳活动异常

地球对太阳辐射的吸收主要在两个区域, 一个是地球表面, 地表吸收可见光而升温, 另一个是平流层顶, 该区域臭氧层吸收太阳紫外线。受太阳紫外线辐射和中间层沉降 NO_x 的影响, 被臭氧层发生波动, 改变了平流层经向温度梯度, 引起平流层纬向风异常。由于波动只能在弱西风传播[6], 因此纬向风异常将改变行星波在平流层的传播条件, 并通过波-流相互作用将异常传递到对流层。太阳活动周期对地球表面的直接影响较难被观测发现, 但在平流层顶区域却能探测到明显的影响, 引起气候变暖。

3.3. 森林火灾

树木燃烧后释放的粉尘和二氧化碳会在全球范围内发生一定影响。然而在寒带地区, 森林被烧毁之后, 可以使地面的积雪被大面积暴露出来, 积雪将射向地球的太阳光反射回宇宙空间, 这反而会导致火灾地区的气候变凉。

这一发现表明, 人们采取很多措施来避免森林火灾减少造成温室效应的主要气体——二氧化碳和甲烷的排放, 可能在无意之中反而使地区的气候变暖了。

过去的 10 年中, 大规模的火灾经常烧毁这些地区的森林, 在一个较长的火灾季节过尽科学家们预测气候会变暖, 而且在以后的几个世纪中火灾会发生得更加频繁。地球吸收太阳光确实会使气候变暖, 但同时, 将热量反射回宇宙空间又会使气候变凉。

对于森林火灾这种突发的自然灾害, 复杂的生态机制可以使其保持相对动态平衡, 但避免人为的故意纵火现象仍然是目前必须实行的措施。

3.4. 火山爆发

对于火山爆发这类自然灾害, 通过采用完全耦合的 CESM-LE, 精选三次的喷发, 即 1963 年的阿贡喷发、1982 年的 El Chichón 喷发和 1991 年的 Pinatubo 喷发。随后从背景噪音中分离出对五次喷发的气候响应, 同时去除每次喷发前五年的平均气候学数据和长期线性趋势。

使用 1982 年 El Chichón 喷发前三年的平均气候学数据。在计算 MME 平均值时, 首先在每个模型中取平均值, 然后得到多峰平均值; 这样, 所有模型都被赋予同等的权重。在不考虑模型偏差的情况下, 直接对所有模型的参数进行平均, 也得到了类似的结果。在火山喷发后, 接下来的三年里都有降温的趋势。

图 11 中显示了五次大型热带火山爆发的全球平均卫星异常及其观测和模拟的合成值, 可以看到曲线整体呈下滑趋势, 考虑到全球冷却暂停和厄尔尼诺[7]反应可能与三月出现的异常高温有关。

重新分析、观测和模拟中显示了火山爆发后第一个冬季全球卫星综合异常(图 12)。

可以发现的是在第一个北半球冬季, 全球降温的暂停似乎是热带地区显著的厄尔尼诺现象升温和欧亚大陆的强烈升温的结果。

由此我们得到的结论是大型火山喷发提供了一种重要的外部强迫, 导致全球地表显著降温, 通常持续 3 年。由于厄尔尼诺现象, 部分火山喷发引起的全球冷却可能不会发生。

综上所述, 新冠疫情短时期会带来温室气体的减排, 如 NO_x 下降高达 30%, 将带来高达 0.01 摄氏度的短期降温, 二氧化碳的减少带来的降温幅度约为 0.01 ± 0.005 摄氏度; 太阳活动周期对地球表面的直

接影响较难被观测发现,但在平流层顶区域却能探测到明显的影响,引起气候变暖;森林火灾确实会使气候变暖,但将热量反射回宇宙空间却会使气候变凉。火山喷发会带来三年的显著降温,但部分收到厄尔尼诺影响的地区受到的影响较小。

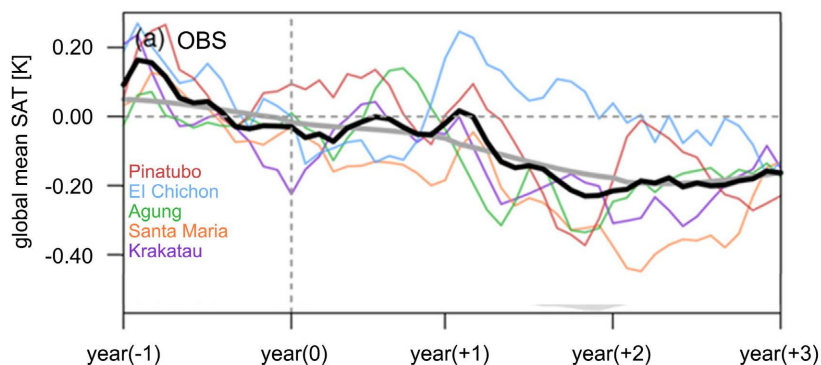


Figure 11. Global mean satellite anomalies of five major tropical volcanic eruptions and their composite values from observations and simulations

图 11. 五次大型热带火山爆发的全球平均卫星异常及其观测和模拟的合成值图

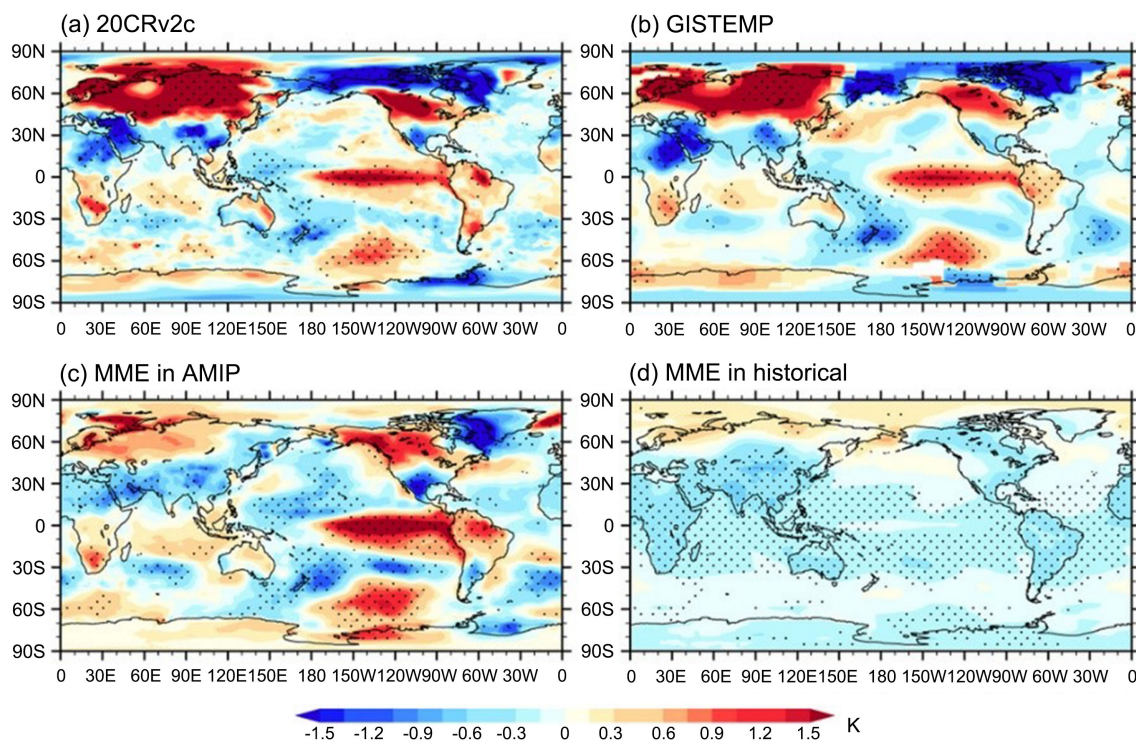


Figure 12. Observed and simulated boreal-winter SAT anomaly after the eruptions

图 12. 观测和模拟了火山喷发后的北冬季 SAT 异常图

4. 影响全球气温变化的主要原因

4.1. 基于 ARIMA 的多元线性回归模型的建立

判断影响全球气温变化的主要原因可以从降水量、海洋表面温度、全球平均温度、CO₂ 浓度、地球吸放热等因素综合考虑。

对未来几年气候的变化进行的预测, 往往可以利用多元线性回归进行求解[8], 首先我们需要剔除一些无关变量对整体模型的影响, 利用逐步回归的方法, 选择对影响气温的显著变量建立模型。

我们假定如下几个变量: CO₂ 浓度、地球的吸热、地球的放热(长波短波因素)、海洋表面温度, 以全球平均温度为因变量, 以如上 4 个解释变量为自变量做多元线性回归。

通过 SPSS 给出相关性检验(表 4)。

Table 4. Correlation test
表 4. 相关性检验

	平均温度	CO ₂ 浓度	长波	短波	海表温度
平均温度	1.000	0.889	-0.712	0.830	-0.744
CO ₂ 浓度	0.889	1.000	-0.352	0.471	-0.571
长波	-0.712	-0.352	1.000	-0.151	0.341
短波	0.830	0.471	-0.151	1.000	-0.477
海表温度	-0.744	-0.571	0.341	-0.477	1.000

通过相关数据可以看出全球与自变量的相关系数大多都在 0.7 以上, 说明所选择变量与全球平均气温高度线性相关[9], 所以这样多元线性回归是合适的。

将 $\mu_0, \mu_1, \mu_2, \mu_3, \mu_4$ 设为相应系数下线性方程:

$$\begin{cases} 13.79 = \mu_0 + 290.7\mu_1 + 61.787\mu_2 + 28.8972\mu_3 + 29.476\mu_4 \\ 13.87 = \mu_0 + 291.2\mu_1 + 65.661\mu_2 + 30.469\mu_3 + 29.741\mu_4 \\ 13.84 = \mu_0 + 290.7\mu_1 + 69.013\mu_2 + 41.562\mu_3 + 31.186\mu_4 \\ \dots \\ 14.62 = \mu_0 + 373.2\mu_1 + 68.332\mu_2 + 51.016\mu_3 + 20.641\mu_4 \end{cases} \quad (3)$$

通过 SPSS 进行的回归分析并求解, 进而可求得相应的多元回归方程为:

$$Y = 12.11052 + 1.354541X_1 + 1.568431X_2 + 3.174641X_3 - 2.6414X_4 \quad (4)$$

对于模型求解可得到了多元线性回归方程, 可知地球放热对于温度影响最大, 其次是 CO₂ 浓度对温度影响较大。

4.2. 基于 ARIMA 的多元线性回归模型的优化与检验

利用我们建立的预测结果较为准确的 ARIMA 时间序列对多元线性回归预测模型进行优化[10], 自回归模型只能适用于预测与自身历史数据相关性较强的数据, 经过我们的检验发现数据符合这一标准, 我们选取(0, 1, 1) ARIMA 模型, 表达式如下

$$y_t = \delta + \sum_{i=1}^p \lambda_i y_{t-i} + \varepsilon_t \quad (5)$$

考虑到可能出现的随机误差, 将模型进行修正

$$y_t = \delta + \sum_{i=1}^p \lambda_i y_{t-i} + \sum_{i=1}^q \psi_i \varepsilon_{t-i} + \varepsilon_t \quad (6)$$

依然将 2020 年作为重要的时间节点, 全球大部分国家的气温都达到了历史新高, 通过 SPSS 软件进行程序实现, 我们的模型也契合了这一预测(图 13, 图 14)。

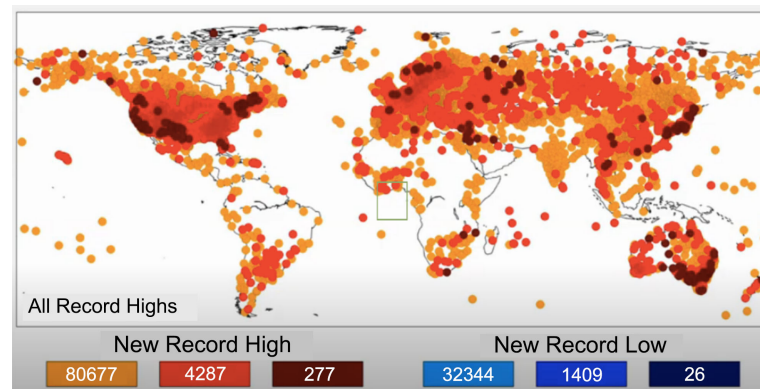


Figure 13. National statistics of 2020 temperature breaking record [11]

图 13. 2020 年气温突破历史记录的国家统计[11]

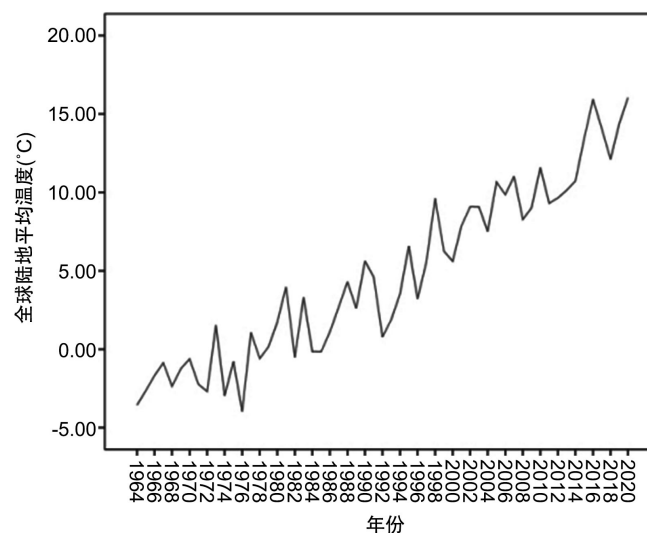


Figure 14. Map of Global Mean Land Temperature from 1964 to 2020 (°C)

图 14. 从 1964 到 2020 全球陆地平均温度图(°C)

我们对数据进行了一阶差分处理

$$(1-L)y_t = \beta + (1+\psi L)\varepsilon_t \quad (7)$$

$$y_t = \beta + y_{t-1} + \varepsilon_t + \psi\varepsilon_{t-1} \quad (8)$$

另外对模型白噪声进行残差检验：通过 SPSS 建模，我们得到残差的 ACF 和 PACF 以及平稳 R2 和对残差进行 Q 检验得到的 P 值，可以看出，所有滞后阶数的自相关系数和偏自相关系数均和 0 没有显著的差异。另外，从平稳 R 方和 1 比较接近，即模型的拟合度较好，准确性较高。

基于 ARIMA (0, 1, 1)模型的多元回归线性模型能够很好地识别该温度数据。

4.3. 基于 ARIMA 的多元线性回归模型的推广

由于最先对模型的假设是在未来一段时期内不会出现新的影响地球气候的人为和非人为因素，并没有测试当特殊因素发生时模型的可行性，故我们以极寒天气的出现对模型进行推广。

极寒天气中较为典型的是极地涡旋气候，它指的是位于地球两极大气层的冷空气和低压，在冬季涡

旋会扩展到北半球, 会将冷空气推向更南方, 在不常有寒冷天气的地区带来温度的急剧下降。

通过 SPSS 系数求解可以得到以下多元回归方程, 当然其中也会包含误差这里用 ε_0 来说明方程如下:
用 Z 来表示极寒天气的变化情况:

$$Z = 3.2897 + 2.5764y_1 + 0.51914y_2 + 1.3611x_1 - 0.5621x_2 - 0.4324x_3 + 0.91x_4 + \varepsilon_0 \quad (9)$$

对于得到的多元回归模型可以分析可得极寒天气 Z 易受全球温度 y_1 变化的影响, 随着全球温度 y_1 的升高, 可得极寒温度 Z 会呈现上升趋势, 极寒天气会受到其它因素的影响, 但是与全球变暖没有矛盾, 基于 ARIMA 的多元回归线性模型也有较大的灵敏度。在极端天气的情况下也同样试用, 应用效果好, 应用场景广。

5. 结论与展望

全球气候变暖已成为国际社会广泛关注的问题, 本文利用 ARIMA 和多元线性回归模型来分析全球的温度变化和影响气温的因素。首先我们对全球平均气温进行可视化与预处理, 得出 2022 年 3 月全球气温的上升是比之前任何十年更大的结论, 我们模型的预测结果显示, 在 2227 年全球平均气温可以达到 20.0033°C。然后我们开发一种机器学习方法对四种自然灾害的影响进行分析, 给出了长期线性趋势, 得出新冠肺炎会带来 0.01°C 左右的短期降温, 火山喷发会带来三年的显著降温, 森林火灾的升温降温效应几乎抵消, 太阳活动异常带来气候变暖的结论。接着对于影响全球气温变化的原因, 我们从降水量、海洋表面温度、全球平均温度、二氧化碳浓度、地球吸放热等因素进行考虑, 利用逐步回归法通过 SPSS 进行相关性检验求解多元线性微分方程, 得出地球放热对温度影响最大的结论。最后我们对模型进行了优化和推广, 将 ARIMA 模型与多元线性回归预测模型进行结合, 利用极寒天气情况对模型进行推广, 检验模型的灵敏度, 最终的拟合效果较好, 应用场景较广。

我们建立的模型能够与实际情况紧密联系, 对于突发的自然灾害、极端天气的灵敏度较高, 只需要内生变量而不需要借助其他外生变量, 具有很好的通用性和推广性。然而, 在数据处理过程中, 我们没有考虑当人类科技获得重大突破时, 对温室气体的改善情况, 并且要求时序数据是稳定的, 或者是通过差分化后是稳定的, 这也是我们下一步优化改进的方向。

对于未来, 我们应控制能源排放, 特别是二氧化碳排放, 及时制定全球化的条约, 达成“只有一个地球”的绿色环保意识。加强发达国家和不发达国家的技术交流和能源转型, 使用多种种类的清洁能源。实现绿色复苏是必要的, 这是我们义不容辞的责任。

参考文献

- [1] 刘政阳, 李挺宇. 中国地表气温上升趋势明显[J]. 生态经济, 2019, 35(6): 9-12.
- [2] 陈振坤, 贾积身, 原晨冉, 徐瑞. 多元非线性回归模型在气候变暖方面的应用——以加拿大地区为例[J]. 河南科技学院学报(自然科学版), 2020, 48(3): 60-67.
- [3] 张宸豪, 冯曦, 冯卫兵, 刘涛, 丁坤. 基于大数据分析下的气候模型[J]. 海洋科学, 2020, 44(10): 1-11.
- [4] National Oceanographic and Atmospheric Administration. NOAA Research News & Features (2022) Carbon Dioxide Now More than 50% Higher than Pre-Industrial Levels. <https://www.noaa.gov/news-release/carbon-dioxide-now-more-than-50-higher-than-pre-industrial-levels> Malik
- [5] Forster, P.M., Forster, H.I., Evans, M.J., et al. (2020) Publisher Correction Current and Future Global Climate Impacts Resulting from COVID-19. *Nature Climate Change*, 10, 971. <https://doi.org/10.1038/s41558-020-0904-z>
- [6] Fang, J.Q. (2020) Study on 11-Year Solar Activity Signal Characteristics in Northern Hemisphere Winter Climate under Global Warming. East China Normal University, Shanghai.
- [7] Dogar, M.M., Hermanson, L., Scaife, A.A., Visoni, D., Zhao, M., Hoteit, I., Graf, H.-F., Dogar, M.A., Almazroui, M. and Fujiwara, M. (2022) A Review of El Niño Southern Oscillation Linkage to Strong Volcanic Eruptions and Post-Volcanic Winter Warming. *Earth Systems and Environment*, 7, 15-42.

<https://doi.org/10.1007/s41748-022-00331-z>

- [8] 彭瑞林, 张玉瑞, 郭瑞伟. 基于多元回归的全球气候变化模式分析[J]. 齐鲁工业大学学报, 2020, 34(4): 61-68.
- [9] Peng, R.L., Zhang, Y.R. and Guo, R.W. (2020) Model Analysis of Global Climate Change Based on Multiple Regression. *Journal of Qilu University of Technology*, **34**, 61-68.
- [10] 王玉华. 基于 ARIMA 模型和 LSTM 神经网络的全球气温预测分析[J]. 科技创新, 2021(35): 166-170.
- [11] 2020 年全球温度记录. 基于 8000 多个气象站 40 年来记录的原始数据、记录高温和记录低温的可视化[EB/OL]. <https://berkeleyearth.org/data>, 2022-11-26.