

针对全球变暖问题的相关数据的建模量化分析

孙一泽, 黄文毅

上海理工大学机械工程学院, 上海

收稿日期: 2023年3月17日; 录用日期: 2023年5月22日; 发布日期: 2023年5月29日

摘要

本文旨在针对自上个世纪开始出现的全球变暖效应, 在进入21世纪后出现了Hiatus (全球变暖停滞状态) 现象, 使得公众开始怀疑全球变暖的真实性这一具体情况, 通过大量数据分析全球变暖产生的原因, 深入挖掘并量化分析极端天气现象的产生原因与气候变化之间所存在的数学关系, 通过具体国家, 选取代表性城市通过历史数据挖掘出温度的时空变化趋势, 从历史数据中找寻温度变化规律。同时本文站在气候尺度上看待全球变暖问题, 考虑到地表温度变化、大气层中二氧化碳浓度变化以及海洋吸热等一系列因素, 量化相关数据, 建立有利于非专业人士理解和认识的全球气候模型和极端天气模型, 对未来气候变化进行预测。

关键词

GM模型, 灰色关联度, 预测气候变化, 极端天气, MATLAB应用

Modeling Quantitative Analysis of Data Related to the Problem of Global Warming

Yize Sun, Wenyi Huang

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

Received: Mar. 17th, 2023; accepted: May 22nd, 2023; published: May 29th, 2023

Abstract

The purpose of this paper is to address the specific situation that Hiatus (global warming hiatus) phenomenon has appeared since the last century, which makes the public begin to doubt the reality of global warming. Through analyzing the causes of global warming with a large number of data, the mathematical relationship between the causes of extreme weather phenomena and climate change is deeply explored and quantitatively analyzed. Through the historical data of specific countries and representative cities, the temporal and spatial trends of temperature are mined,

and the rules of temperature change are found from the historical data. At the same time, this paper looks at global warming from the climate scale, taking into account a series of factors such as surface temperature change, carbon dioxide concentration change in the atmosphere and ocean heat absorption, quantifying relevant data and establishing global climate models and extreme weather models that are conducive to non-professionals' understanding and cognition, so as to predict future climate change.

Keywords

GM Model, Grey Correlation Degree, Predicting Climate Change, Extreme Weather, MATLAB Application

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

从工业革命开始, 由于大气中的二氧化碳浓度不断上升, 温室效应的效果也随之增强。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, 不仅会危害自然生态系统的平衡, 还威胁人类的生存。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°C , 几乎未变化, 这种现象叫做 Hiatus (全球变暖停滞状态)。正因为出现全球变暖停滞现象, 使公众对全球变暖产生了怀疑。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。导致分歧的原因在于观察问题的角度和范围。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。

吕金莹等[1]运用差分整合移动平均自回归模型(ARIMA)对松嫩平原未来 10 年活动积温变化进行了预测。周宏伟等[2]运用自回归移动平均模型(ARMA)对北京月平均相对温度进行了预测。薛宇峰和杨超梅[3]利用墨西哥帽小波函数分析了近百年来全球气温的变化特征, 并用气温距平序列的 5a 滑动平均计算了信噪比(S/N), 验证了气候突变点的位置, 同时, 利用人工神经网络对全球气温变化趋势进行了预测。彭润龙等[4]综合分析了全球变暖的气候变化问题对全球变暖背景下海洋表面温度的规律变化, 对未来 25 年的气候变化进行了预测, 并就极端天气现象与气候变化的影响因素建立了多元回归分析模型进行研究。

一般系统理论只能建立差分模型, 不能建立微分模型。针对上述问题, 本文采用了灰色理论[5]的微分方程性模型称为 GM 模型。差分模型是一种递推模型, 只能按阶段分析系统的发展, 只能用于短期分析, 只能了解系统显露的变化。而灰色理论, 基于关联度收敛原理、生成数、灰导数、灰微分方程等观点和方法建立了微分方程型模型。这样建立的模型, 适用性强, 计算方便, 结果精确, 可以实现对杂乱无章的原始数据的处理。加拿大是世界上海岸线最长的国家, 气候情况复杂。本文选取五个具有气候代表性的城市, 对其近 35 年的温度变化, 降雨和降雪量进行数据分析, 挖掘出加拿大温度时空变化趋势。观察其温度周期波动情况, 并建立 GM 模型预测未来 25 年的气候变化。

2. 基本假设和符号说明

2.1. 基本假设

- 假设一：针对本文官网提供的数据均为真实有效的数据；
 假设二：地球以外的因素不会对所建立的地球气候模型产生影响；
 假设三：各种社会因素处于稳定状态。假设地球整体环境处于稳定状态；
 假设四：假设加拿大的年平均温度变化趋势与全球表面温度变化趋势一致。

2.2. 符号说明

本文用到的符号以及其意义如表 1 所示。

Table 1. Description of some symbols used in this article

表 1. 本文用到的部分符号说明

符号	说明
m	指标的个数
n	被评价对象的个数
ρ	分辨系数
$\eta_i(k)$	关联系数
W_k	各指标权重
α	发展灰数
μ	内生控制灰数
c	方差比
d	小误差概率
Q	相对误差

3. 模型建立及求解

3.1. 大气二氧化碳含量数据

大气中能产生温室效应的气体近 30 种，其中二氧化碳起重要的作用，从长期气候数据比较来看，在气温和二氧化碳之间存在显著的相关关系。本文采用世界气象组织发布的温室气体公报中全球二氧化碳浓度进行分析(表 2)。

Table 2. Global average carbon dioxide concentration

表 2. 全球二氧化碳平均浓度

年份	CO ₂ 平均浓度/pm	年份	CO ₂ 平均浓度/pm	年份	CO ₂ 平均浓度/pm	年份	CO ₂ 平均浓度/pm
1959	315.98	1975	331.08	1991	355.48	2007	383.71
1960	316.91	1976	332.05	1992	356.27	2008	385.57
1961	317.64	1977	333.78	1993	356.95	2009	387.35
1962	318.45	1978	335.41	1994	358.64	2010	389.85

Continued

1963	318.99	1979	336.78	1995	360.62	2011	391.63
1964	319.62	1980	338.68	1996	362.36	2012	393.82
1965	320.04	1981	340.11	1997	363.47	2013	396.00
1966	321.38	1982	341.22	1998	366.50	2014	398.65
1967	322.16	1983	342.84	1999	368.14	2015	400.00
1968	323.04	1984	344.41	2000	369.40	2016	403.30
1969	324.62	1985	345.87	2001	371.07	2017	405.50
1970	325.68	1986	347.19	2002	373.17	2018	408.52
1971	326.32	1987	348.98	2003	375.78	2019	415.00
1972	327.45	1988	351.45	2004	377.52		
1973	329.68	1989	352.90	2005	379.76		
1974	330.17	1990	354.16	2006	381.85		

3.2. 模型建立和求解

采用 GM(1,n)模型, 设时间序列 $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n)\}$ 有 n 个观察值, 通过累加生成新序列 $X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n)\}$ 则 GM(1,n)模型相应的白化形式微分方程为:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = \mu \tag{1}$$

其中: a 称为发展灰数; μ 称为内生控制灰数。

记

$$B = \begin{pmatrix} -\frac{1}{2}[X^{(1)}(1)+X^{(1)}(2)] & 1 \\ -\frac{1}{2}[X^{(1)}(2)+X^{(1)}(3)] & 1 \\ \dots & \dots \\ -\frac{1}{2}[X^{(1)}(n-1)+X^{(1)}(n)] & 1 \end{pmatrix}$$

$$Y_n = \begin{pmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \dots \\ X^{(0)}(n) \end{pmatrix}$$

称 $Z^{(1)}$ 为 $X^{(1)}$ 的紧邻均值生成序列

$$Z^{(1)} = \{Z^{(1)}(2), Z^{(1)}(3), \dots, Z^{(1)}(n)\}$$

$$Z^{(1)}(k) = \frac{1}{2}(X^{(1)}(k-1) + X^{(1)}(k))$$

则

$$B = \begin{pmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ \dots & \dots \\ -Z^{(1)}(n) & 1 \end{pmatrix}$$

设 $\hat{\alpha}$ 为待估参数向量, $\hat{\alpha} = \begin{pmatrix} a \\ \mu \end{pmatrix}$

则微分方程可表示为

$$Y = B\hat{\alpha} \quad (2)$$

可利用最小二乘法求解, 解得:

$$\hat{\alpha} = (B^T B)^{-1} B^T Y_n \quad (3)$$

求解微分方程, 即可得预测模型:

$$\hat{X}^{(1)}(k+1) = \left[X^{(0)}(1) - \frac{\mu}{a} \right] e^{-ak} + \frac{\mu}{a}, \quad k = 0, 1, 2, \dots, n. \quad (4)$$

对其做累减还原, 即可得到原始数列的灰色预测模型为:

$$\hat{X}^{(0)}(k) = \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k) \quad (5)$$

由灰色预测方法原理, $-a$ 主要控制系统发展态势的大小, 即反映预测的发展态势, 被称为发展系数; μ 的大小反映了数据变化的关系, 被称为灰色作用量, 其中:

- ①当 $-a < 0.3$ 时, GM(1,1)模型可用于中长期预测;
- ②当 $0.3 < -a < 0.5$ 时, GM(1,1)模型可用于短期预测, 中长期预测慎用;
- ③当 $0.5 < -a < 1$ 时, 应采用 GM(1,1)改进模型, 包括 GM(1,1)残差修正模型;
- ④当 $-a > 1$ 时, 不宜采用 GM(1,1)模型, 可考虑其他预测方法。

模型检验一般有残差检验、关联度检验和后验差检验。

1) 残差检验

按预测模型计算 $\hat{X}^{(1)}(i)$, 将 $\hat{X}^{(1)}(i)$ 累减生成 $\hat{X}^{(0)}(i)$, 然后计算原始数列 $X^{(0)}(i)$ 与 $\hat{X}^{(0)}(i)$ 的绝对误差序列及相对误差序列。

$$\Delta^{(0)}(i) = |X^{(0)}(i) - \hat{X}^{(0)}(i)|, \quad i = 1, 2, \dots, n \quad (6)$$

$$\Phi(i) = \frac{\Delta^{(0)}(i)}{X^{(0)}(i)} \times 100\%, \quad i = 1, 2, \dots, n \quad (7)$$

给定 a , 当 $\bar{\Phi} < a$ 成立, 称模型为残差合格模型

a 取 0.01、0.05、0.1 分别为: 优、合格、勉强合格

2) 关联度检验

根据前面所述关联度的计算方法算出 $\hat{X}^{(0)}(i)$ 与原始序列 $X^{(0)}(i)$ 的关联系数, 然后计算出关联度, 根据经验, 当 $\rho = 0.5$ 时, 关联度大于 0.6 便满意了。

设原始数据序列 $X_0^{(0)} = \{x_0^{(0)}(1), x_0^{(0)}(2), \dots, x_0^{(0)}(n)\}$ 为参考序列, 用 m 种灰色建模得模型值分别为 $X_i^{(0)} = \{x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(n)\}$, $i = 1, 2, \dots, m$, 求出该点 m 个序列与参考序列的邓氏关联度 $\varepsilon_i (i = 1, 2, \dots, m)$, 如果 $\gamma_i (i = 1, 2, \dots, m)$ 在所有关联度中最大, 则第 i 种灰色建模方法为所建模型中最好的

模型。

3) 后验差检验

a) 计算原始序列标准差:

$$S_1 = \sqrt{\frac{\sum [\mathbf{X}^{(0)}(i) - \bar{\mathbf{X}}^{(0)}]^2}{N}} \quad (8)$$

b) 计算绝对误差序列的标准差:

$$S_2 = \sqrt{\frac{\sum [\Delta^{(0)}(i) - \bar{\Delta}^{(0)}]^2}{N}} \quad (9)$$

c) 计算方差比:

$$C = \frac{S_2}{S_1} \quad (10)$$

d) 计算小误差概率:

$$P = P\{|\Delta^{(0)}(i) - \bar{\Delta}^{(0)}| < 0.6745S_1\} \quad (11)$$

令

$$e_i = |\Delta^{(0)}(i) - \bar{\Delta}^{(0)}|, S_0 = 0.6745S_1$$

则

$$P = P\{e_i < S_0\} \quad (12)$$

若相关误差, 关联度、后验差检验在允许范围之内, 则可用所建模型进行预测(表 3)。

Table 3. Accuracy class

表 3. 精度等级

P	C	精度
>0.95	<0.35	一级
>0.80	<0.50	二级
>0.70	<0.65	三级
≤0.70	≤0.70	四级

3.3. 灰色关联度分析介绍

灰色系统的应用范畴大致分为: 灰色关联分析、灰色预测、灰色决策、灰色预测控制。灰色系统理论是人们认识客观系统改造客观系统的一个新型的理论工具。灰色关联分析有别于过去主要采取回归分析、方差分析, 主成分分析的因素分析方法, 灰色系统理论采用的关联分析方法可以克服过去算法需要大量数据作为基础, 计算量大的弊端。

利用灰色关联分析进行综合评价的步骤是:

1) 确定评价指标体系, 收集评价数据。

设 n 个数据序列形成如下矩阵:

$$(\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_n) = \begin{pmatrix} x'_1(1) & x'_2(1) & \cdots & x'_n(1) \\ x'_1(2) & x'_2(2) & \cdots & x'_n(2) \\ \vdots & \vdots & \vdots & \vdots \\ x'_1(m) & x'_2(m) & \cdots & x'_n(m) \end{pmatrix}$$

其中 m 为指标的个数,

$$\mathbf{X}'_i = (x'_i(1), x'_i(2), \dots, x'_i(m))^T, \quad i = 1, 2, \dots, n.$$

2) 确定参考数据列

参考数据列应该是一个理想的比较标准, 可以以各指标的最优值(或最劣值)构成参考数据列, 也可根据评价目的选择其它参照值。记作

$$\mathbf{X}'_0 = (x'_0(1), x'_0(2), \dots, x'_0(m))^T$$

3) 对指标数据进行无量纲化

无量纲化后的数据序列形成如下矩阵:

$$(\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_n) = \begin{pmatrix} x_0(1) & x_1(1) & \cdots & x_n(1) \\ x_0(2) & x_1(2) & \cdots & x_n(2) \\ \vdots & \vdots & \vdots & \vdots \\ x_0(m) & x_1(m) & \cdots & x_n(m) \end{pmatrix}$$

常用的无量纲化方法有均值化法、初值化法和 $\frac{x - \bar{x}}{S}$ 标准化变换等。

均值化法:

$$x_i(k) = \frac{x'_i(k)}{\frac{1}{m} \sum_{k=1}^m x'_i(k)}, \quad i = 0, 1, \dots, n; \quad k = 1, 2, \dots, m. \quad (13)$$

初值化法:

$$x_i(k) = \frac{x'_i(k)}{x'_i(1)}, \quad i = 0, 1, \dots, n; \quad k = 1, 2, \dots, m. \quad (14)$$

或采用内插法使各指标数据取值范围(或数量级)相同。

4) 逐个计算每个被评价对象指标序列(比较序列)与参考序列对应元素的绝对差值

$$|x_0(k) - x_i(k)| \quad (k = 1, 2, \dots, m; \quad i = 1, 2, \dots, n),$$

n 为被评价对象的个数。

5) 确定两级最小差与两级最大差

两级最小差

$$\min_{i=1}^n \min_{k=1}^m |x_0(k) - x_i(k)|$$

两级最大差

$$\max_{i=1}^n \max_{k=1}^m |x_0(k) - x_i(k)|$$

6) 计算关联系数

$$\eta_i(k) = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|}, k = 1, 2, \dots, m.$$

式中 ρ 为分辨系数, 在 $(0, 1)$ 内取值, 若 ρ 越小, 关联系数间差异越大, 区分能力越强, 常 ρ 取 0.5。

如果 $\{x_0(k)\}$ 为最优值数据列, $\eta_i(k)$ 越大, 越好;

如果 $\{x_0(k)\}$ 为最劣值数据列, $\eta_i(k)$ 越大, 越不好。

7) 计算关联度

对各评价对象(比较序列)分别计算其个指标与参考序列对应元素的关联系数的均值, 以反映各评价对象与参考序列的关联关系, 并称其为关联度, 记为:

$$\gamma_i = \frac{1}{m} \sum_{k=1}^m \eta_i(k) \tag{15}$$

8) 如果各指标在综合评价中所起的作用不同, 可对关联系数求加权平均值即

$$\gamma'_i = \frac{1}{m} \sum_{k=1}^m W_k \cdot \eta_i(k), k = 1, 2, \dots, m. \tag{16}$$

式中 W_k 为各指标权重。

9) 依据各观察对象的关联度, 进行排序, 得出综合评价结果。

4. 模型求解

4.1. 求解结果分析

利用上述 GM 模型, 通过现有全球二氧化碳平均浓度数据以及温度变化情况对未来 25 年不同因素影响下的温度变化进行了预测。

图 1(a)为二氧化碳浓度的变化情况。小概率误差 $P = 1$, 方差比 $C = 0.0045$, 相对误差 $Q = 1.0399e^{-5}$, 通过精度比较我们发现小概率误差 $P = 1 > 0.95$, 方差比 $C = 0.0045 < 0.35$, 相对误差 $Q = 1.0399e^{-5} < 0.95$, 模型精度达到一级。由此可见 GM 模型可适用于预测温度变化[6]。图 1(b)为海洋温度和二氧化碳浓度两个因素影响情况下温度变化的情况。小概率误差 $P = 0.8143$, 方差比 $C = 0.3937$, 相对误差 $Q = 0.06024$, 通过精度比较发现小概率误差 $P = 0.8143 > 0.80$, 方差比 $C = 0.3937 < 0.50$, 相对误差 $Q = 0.06024 < 0.95$, 模型精度较高。图 1(c)为海洋温度, 全球二氧化碳浓度以及降雨量和降雪量多个因素的影响下温度变化的情况。可以看出 GM 模型预测精度等级会降低, GM 模型不能像神经网络那样考虑多因素问题, 但是它可以作为一个简单模型来预测未来 25 年的变化, 有利于非专业的简单理解。通过对比可以发现二氧化碳浓度对全球的温度变化具有显著的影响, 并且全球的温度变化依旧处于一个上升趋势。

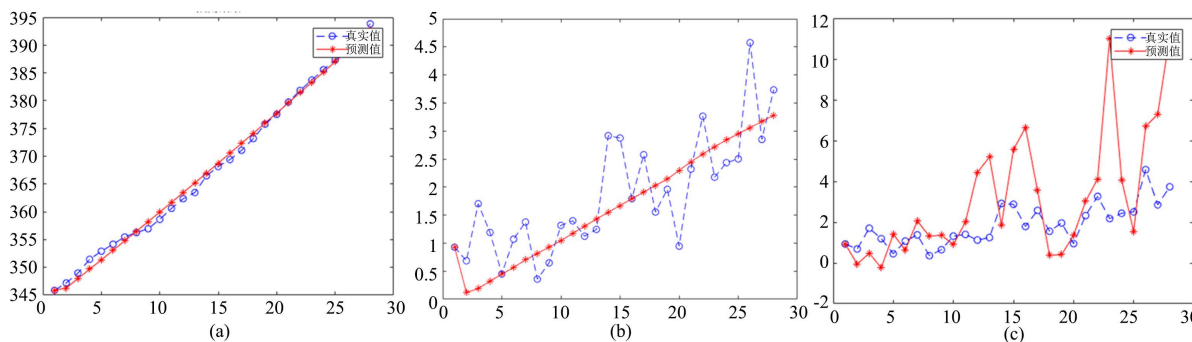


Figure 1. Temperature change under the influence of different factors

图 1. 不同因素影响下温度的变化

4.2. 结果检验分析

通过多因素耦合神经网络 RBF 对 GM 模型预测结果进行检验。

如图 2 所示红色的为实际对应的公式, 蓝色为输入因素得到的预测结果, 从拟合度计算结果来看, $R^2 = 0.98$, 均方根误差 $RMSE = 13.74$, 平均相对误差为 0.1185, 预测精度较高。

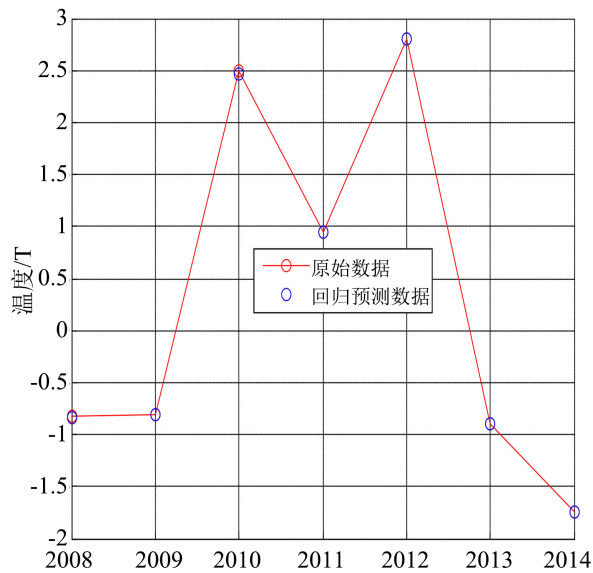


Figure 2. Neural network regression prediction data

图 2. 神经网络回归预测数据

如图 3, 由后续 25 年预测结果可以看出神经网络预测的结果基本与 GM 预测趋势一致, 在海洋温度, 全球二氧化碳浓度以及降雨量和降雪量多个因素影响下, 全球温度将持续走高。

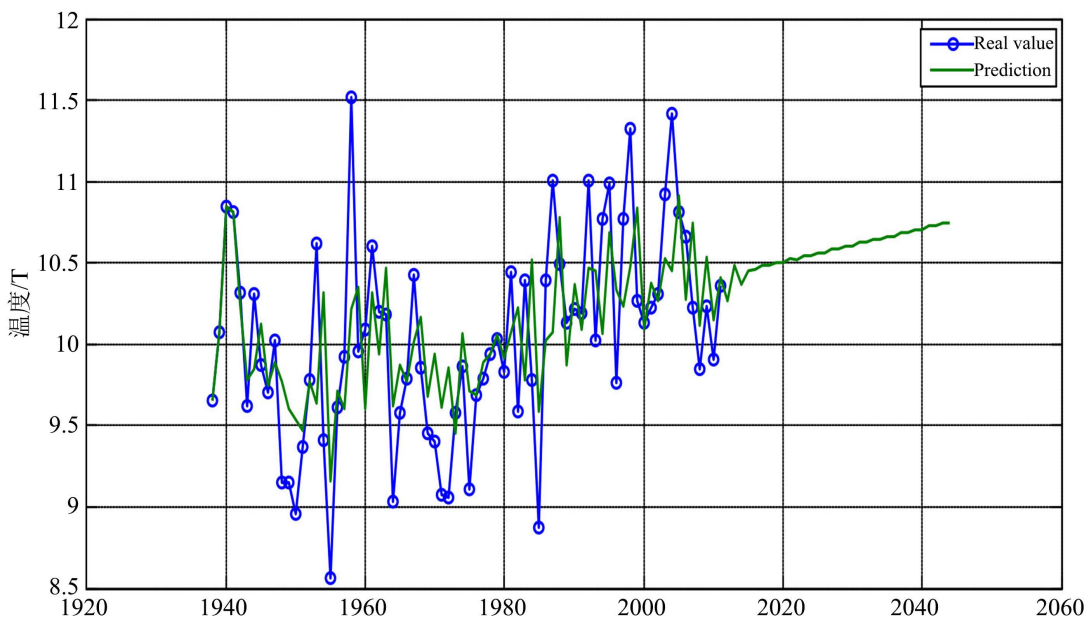


Figure 3. Forecast of temperature from 1938 to 2044

图 3. 1938~2044 年气温预测

5. 结论

从海温图中提取与经纬度有关的海温数据, 收集全球二氧化碳浓度变化, 将海温, 加拿大五个站点的平均气温, 全球二氧化碳的浓度作为变化要素从而建立简化的模型去预测未来 25 年的气候变化, 所建立的模型区别于复杂的专业气候模型。在本文中采用 GM 模型通过单因素分析, 多因素分析来观察模型的精度, 在影响因素较少的情况下预测精度较高, 在综合因素较多的情况下预测精度较低。最后采用神经网络 RBF 对 GM 模型预测的结果进行对比, 验证得到预测结果基本一致。

参考文献

- [1] 吕金莹, 闫超, 贾天宇, 王芳媛, 孙洪超, 董守坤, 龚振平. 松嫩平原活动积温变化及其对作物产量的影响[J]. 生态学杂志, 2019, 38(11): 3349-3356.
- [2] 周宏伟, 张梦晓, 高峰云, 等. 基于 ARIMA 模型对北京市月平均相对湿度的预测[J]. 考试周刊, 2017(59): 106-108.
- [3] 薛宇峰, 杨超梅. 近百年全球气温变化及其趋势预测[J]. 四川气象, 2006(3): 16-19.
- [4] 彭润龙, 张亚如, 郭荣伟. 基于多元回归全球气候变化的模型分析[J]. 齐鲁工业大学学报, 2020, 34(4): 61-68.
- [5] 唐丽芳, 贾冬青, 孟庆鹏. 用 MATLAB 实现灰色预测 GM(1,1)模型[J]. 沧州师范专科学校学报, 2008, 24(2): 35-37.
- [6] 黄振宸. 基于 ARIMA 模型和 GM 模型的全球变暖与台风强度的相关性分析[J]. 科技风, 2018(6): 119-124.