安徽农业大学学报, 2012, 39(5): 837-842
Journal of Anhui Agricultural University
网络出版时间：2012-8-27 11:59:54
[URL] http://www.cnki.net/kcms/detail/34.1162.S.20120827.1159.015.html

基于季节 ARIMA 模型的铜陵市气温序列的预报

沈 艳[1]，张庆国[2*]，叶静芸[2]
(1. 安徽农业大学信息与计算机学院，合肥 230036；2. 安徽农业大学理学院，合肥 230036)

摘 要：运用 EVIEWS 软件，对铜陵市 48 年来的月平均气温时间序列进行统计分析，并对该动态数据进行建模和预测。采用差分方法对样本数据进行预处理，然后定阶，并进行参数估计，建立季节 ARIMA 模型对铜陵市气温数据进行预报。预报结果显示，季节 ARIMA 模型的平均绝对误差值为 0.875。将 ARIMA 模型预报结果与径向基（radial basis function, RBF）神经网络模型的预报值比较可知，其预报结果优于 RBF 神经网络的预测结果。

关键词：时间序列；ARIMA 模型；月平均气温；铜陵市
中图分类号：P468.021 **文献标识码**：A **文章编号**：1672−352X (2012)05−0837−06

Prediction of temperature time series of Tongling city based on season ARIMA model

SHEN Yan[1], ZHANG Qing-guo[2], YE Jing-yun[2]
(1. School of Science, Anhui Agricultural University, Hefei 230036;
2. School of Information & Computer, Anhui Agricultural University, Hefei 230036)

Abstract: In this article, we analyzed the time series data of month mean temperature in Tongling city using EVIEWS software, and got modeling prediction according to the dynamical data. We preprocessed the sample data using difference method, and then made sure the model order and estimated the parameter values for establishing the season autoregressive integrated moving average (ARIMA) model to fit the time series. The prediction results showed that the average absolute error of the season ARIMA model is 0.875. Comparing the result of ARIMA model and RBF（radial basis function）neural network, the season ARIMA model is better than radial basis function (RBF) neural network.

Key words: time series; ARIMA model; month mean temperature; Tongling city

时间序列，又称动态数据，是指随时间的顺序记录的一列有序数据。20 世纪 60 年代，美国学者 Box 和英国统计学者 Jenkins 提出了一整套关于时间序列分析、预报和控制的方法，被称为 Box-Jenkins 建模方法[1-2]。时间序列建模的方法很多，求和自回归移动平均（ARIMA）模型就是其中之一。

掌握天气变化趋势对于天气预报具有重要意义。人们根据可靠的天气预报，就能够适时地安排生产和生活，为国民经济建设服务，更能够减少气象灾害的损失。因此，可以说，准确及时的天气预报有利于国家的经济建设和国防建设，在保障人民生命财产安全、生活条件等方面具有极大的社会效益和经济效益。

本研究收集了 1960 年～2008 年（共 49 年）铜陵市[3]的月平均气温的动态数据，先是对月平均温度序列分析，观察其变化规律，建立季节模型，然后再对该序列进行模型的识别、参数估计、模型检验以及基于该模型的预测。在实际生产和生活中，对于温度的研究和预测具有很大的意义。

1 原理与方法

时间序列分析是处理动态数据的一种比较有效的时域分析方法，通过观察动态数据的变化规律，继而对数据进行控制并预测时间序列的未来变化趋势[2,4]。

收稿日期：2012-03-29
基金项目：国家自然科学基金项目（70271062，40771117）和安徽省级重点科研基金项目（KJ2010A121）共同资助。
作者简介：沈 艳，女，硕士研究生。
* **通讯作者**：张庆国，男，博士，教授。E-mail：qgzhang@ahau.edu.cn

1.1 ARIMA 模型[2,5-7]

求和自回归移动平均（auto regressive integrated moving average，ARIMA）模型拟合的是差分平稳序列，实际上就是差分运算和 ARMA 模型的结合。对于含有非季节性的时间序列进行建模，可以使用 ARIMA（p,d,q）模型。它可以通过适当的 d 阶（d 为整数）差分运算使序列平稳。设 X=（$x_1,x_2,\cdots,x_t,\cdots,x_n$）为一个等时间间隔为 t 的时间序列，ε_t 为误差序列。ARIMA 模型的一般形式为：

$$\Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t$$

其中，$\Phi(B)=1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p$ 为平稳可逆 ARMA(p, q)模型的自回归系数多项式；

$\Theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$ 为平稳可逆 ARMA(p, q)模型移动平均系数多项式；

x_t 为非平稳序列，d 为差分运算的阶数，ε_t 为零均值白噪声序列。

另外，ARMA(p, q)模型就是当 d=0 时的 ARIMA 模型的特殊形式。

1.2 ARIMA(p, d, q)$_s$ 季节模型

季节性是指在某一个固定的时间间隔上，序列的某种特征重复出现，水文时间序列过程具有较为明显的准周期性变化。关于季节性周期的设定，一般月际资料的水文时间序列的季节周期为 12 个月；季度资料的时间序列的季节周期为 4 个季。

对于含有明显季节因素的时间序列（如本文的月平均温度），ARIMA 模型也可以对其建模。简单的季节模型通常通过以周期为步长的差分运算就可以将该时间序列中的季节信息提取充分，使该时序变成平稳序列，它的残差序列也是一平稳序列。ARIMA 季节模型[8]的一般形式为：

$$\Phi(B)\nabla_S^D\nabla^d x_t = \Theta(B)\varepsilon_t$$

其中，S 为周期步长，d 为提取趋势信息用的差分阶数。

所以，对于不含有趋势效应的时间序列，季节性 ARIMA 模型实际上就是对其进行季节差分，以便提取时序中的季节信息。但是，在实际操作中，能够提取足够过的信息量就可以了，对于时间序列的差分运算要防止过分差分的现象。

1.3 ADF 检验

ADF 检验[2,9]（Augmented Dickey-Fuller）是对 DF 检验进一步的修正，使之能够适用于多阶自回归的平稳性检验，即假设数据生成过程服从有单位根的 p 阶自回归过程，即 AR(p)过程为：

$$x_t=\phi_1 x_{t-1}+\phi_2 x_{t-2}+\cdots+\phi_p x_{t-p}+\varepsilon_t$$

其中：ε_t 独立分布，令 $\rho=\phi_1+\phi_2+\cdots+\phi_p-1$。

若序列 x_t 平稳，则 $\phi_1+\phi_2+\cdots+\phi_p<1$，等价为 $\rho<0$；若序列 x_t 非平稳，则至少存在一个单位根，使得 $\phi_1+\phi_2+\cdots+\phi_p=1$，即为 $\rho=0$。则 AR(p)自回归过程的单位根检验的假设条件可以如下确定：

H_0：ρ =0（序列 x_t 非平稳）

H_1：ρ ＜0（序列 x_t 平稳）

构造 ADF 检验统计量：

$\tau=\dfrac{\hat{\rho}}{S(\hat{\rho})}$，其中，$S(\hat{\rho})$ 为参数 ρ 的样本标准差。

这样，就可以通过蒙特卡洛方法得到 τ 统计量的临界值表。

1.4 建模步骤

使用 ARIMA 模型建模一般分为 3 个阶段：模型识别、参数估计[10]和模型检验。建立 ARIMA 模型时要求预处理后的时间序列是平稳的[11]，所以在建立 ARIMA 模型之前必须对时间序列进行平稳性检验，以保证预处理后的数据具有平稳性。对处理后时间序列用上述 3 个步骤反复调用，直到得到用于预测的最优模型为止，其图 1 即为 ARIMA 模型应用的一般步骤。下面通过实例分析，对于具有季节效应的时间序列资料进行分析和预测。以下结果均在 EVIEWS 5.0 软件运行下分析所得。

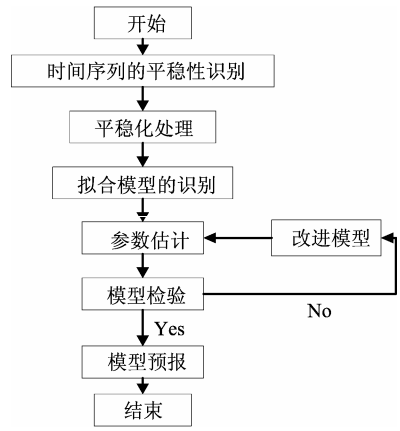


图 1 ARIMA 模型应用的一般步骤

Figure 1 The common steps of ARIMA model application

2 实例应用

2.1 数据预处理[12]及特征描述

以下数据为铜陵市 1960 年～2008 年历年月平均气温（共 586 个，缺失 2 个）。对于缺失值采用相邻两点的均值替代。绘制的时序图如图 2 所示，其中横坐标为年份，纵坐标为月平均气温的温度值。

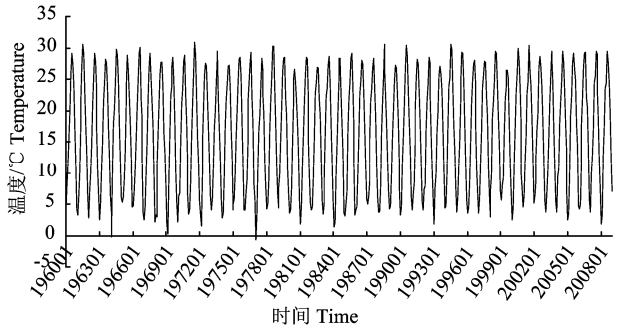


图 2　气温时序

Figure 2　The sequence of temperatures

从气温的时序图上可以看出序列并无明显的趋势性，相对来说序列比较平稳，但是具有很强的周期性。图 3 为时间序列各成分的分解：（1）气温序列中的不规则波动成分；（2）气温序列季节调整后的序列；（3）气温序列的季节因子序列部分；（4）气温序列的趋势与循环波动叠加后的部分。从图 3 可以看出，该时间序列有个比较明显的有规律的波动，季节波动。图 3（4）中可以看出近几十年的气温变化表现为先略微下降，到 90 年代后又开始微微上升，其原因可能是因为工业的发展，但总体上变化幅度不大。

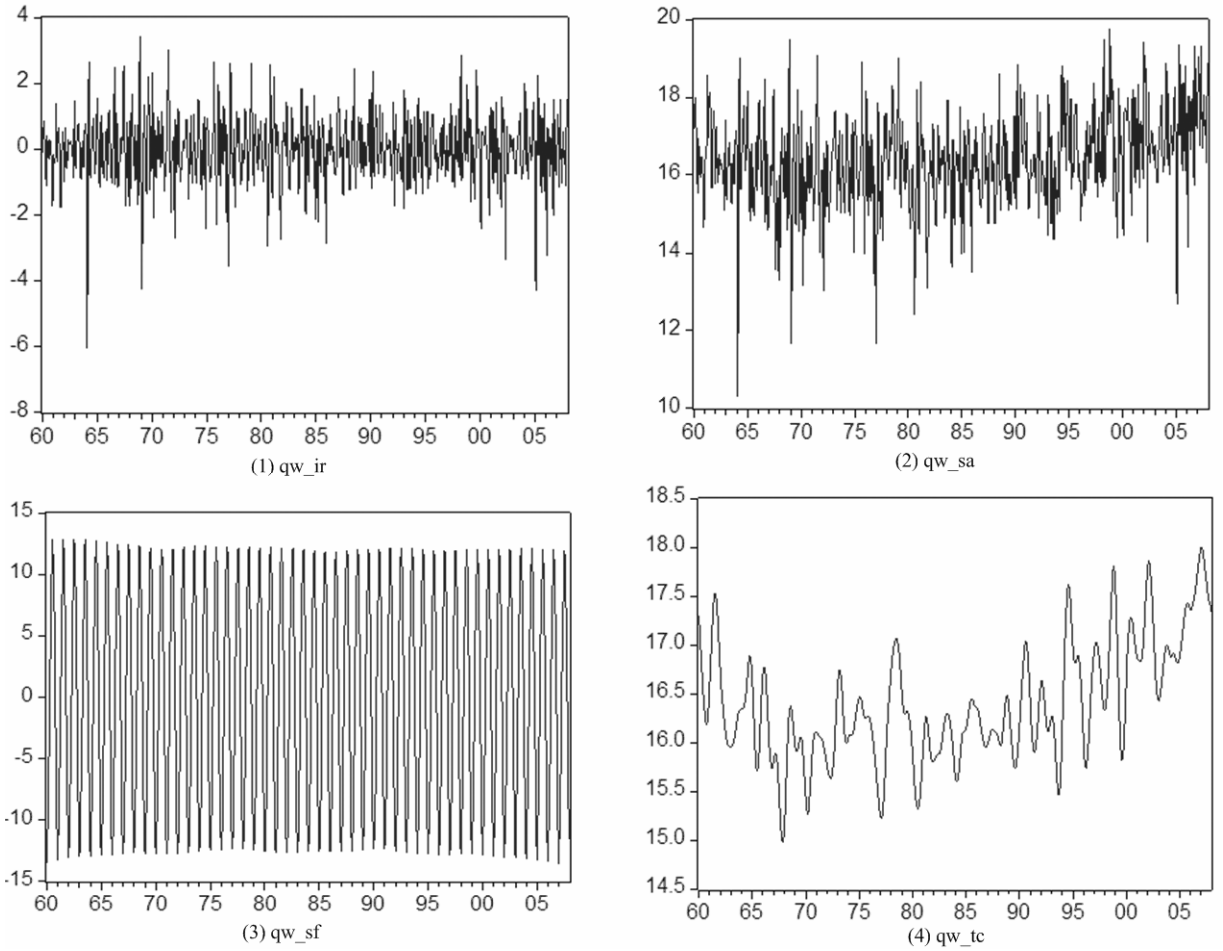


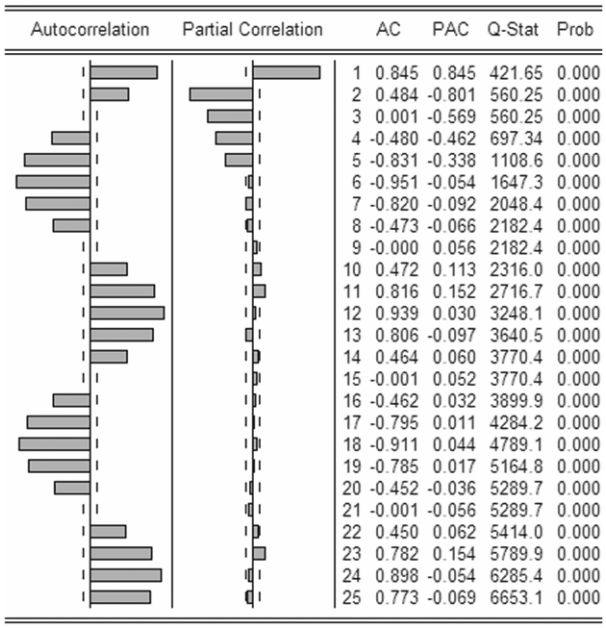
图 3　气温时序的各部分波动因素分解

Figure 3　The different undulatory factors of temperature series

从图 3 和图 4（月平均气温序列的自相关和偏自相关）可以看出，铜陵地区的月平均气温时间序列没有较强的趋势性，但是具有很强的周期性，蕴含着以年为固定周期的周期性。通过消除数据的周期性，计算其自相关函数和偏自相关函数，以尝试用季节性 ARIMA 模型拟合时间序列。为了较好的提取数据的周期信息，对序列进行 12 步的周期差分，以便提取季节波动信息。设数据变换后的差分序列为 y_t，则：$y_t = \nabla_{12} = x_t - x_{t-12}$。

对月平均气温序列进行 12 步周期差分，提取季节效应后的序列时序如图 5 所示，且时序图显示差分后序列类似平稳。差分后的 12 阶处仍有一定的显著性，可能还包含一定的周期性因素，但是根据对差分后序列再进行差分，其结果显示为序列显著不

平稳，故只进行一次 12 步差分即可，以防止序列的过差分。再对差分后序列进行单位根检验，如图 6。结果表明 ADF 值为-3.9206，比 EVIEWS 给出的显著性水平 1%～10%的 ADF 临界值都小，所以拒绝原假设，即序列不存在单位根，进一步说明季节差分后的序列平稳。



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.845	0.845	421.65	0.000
		2	0.484	-0.801	560.25	0.000
		3	0.001	-0.569	560.25	0.000
		4	-0.480	-0.462	697.34	0.000
		5	-0.831	-0.338	1108.6	0.000
		6	-0.951	-0.054	1647.3	0.000
		7	-0.820	-0.092	2048.4	0.000
		8	-0.473	-0.066	2182.4	0.000
		9	-0.000	0.056	2182.4	0.000
		10	0.472	0.113	2316.0	0.000
		11	0.816	0.152	2716.7	0.000
		12	0.939	0.030	3248.1	0.000
		13	0.806	-0.097	3640.5	0.000
		14	0.464	0.060	3770.4	0.000
		15	-0.001	0.052	3770.4	0.000
		16	-0.462	0.032	3899.9	0.000
		17	-0.795	0.011	4284.2	0.000
		18	-0.911	0.044	4789.1	0.000
		19	-0.785	0.017	5164.8	0.000
		20	-0.452	-0.036	5289.7	0.000
		21	-0.001	-0.056	5289.7	0.000
		22	0.450	0.062	5414.0	0.000
		23	0.782	0.154	5789.9	0.000
		24	0.898	-0.054	6285.4	0.000
		25	0.773	-0.069	6653.1	0.000

图 4 自相关和偏自相关结果

Figure 4 The results of ACF and PACF

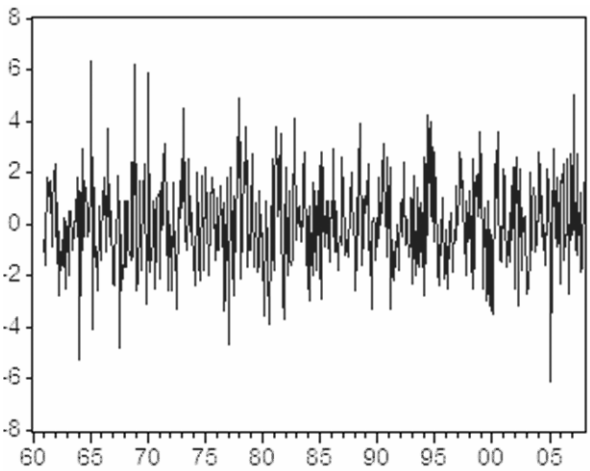


图 5 月度气温 12 步差分后时序

Figure 5 The time series of month temperature after 12 steps difference

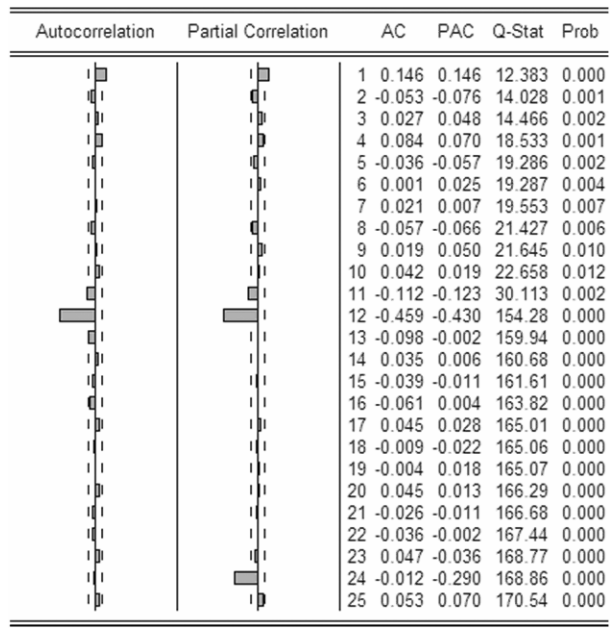
2.2 模型的识别及参数估计

模型的模式识别[13]主要是通过对铜陵市月平均气温的历史数据进行差分处理后，对序列的自相关、偏自相关进行观察得到的，其季节长度 S 可由实际背景分析得到，本文中的季节长度 S 为 12。如果样本的自相关、偏自相关图既不截尾也不拖尾，而且不是呈线性衰减趋势，相反地，在相应于周期 S 的整倍数点上，自相关（或偏自相关）函数出现绝对值相当大的峰值并呈现振荡变化，就可以判断数据序列适用于乘积季节模型。

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.920600	0.0020
Test critical values:	1% level	-3.441553	
	5% level	-2.866374	
	10% level	-2.569404	

图 6 单位根检验结果

Figure 6 The result of ADF



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.146	0.146	12.383	0.000
		2	-0.053	-0.076	14.028	0.001
		3	0.027	0.048	14.466	0.002
		4	0.084	0.070	18.533	0.001
		5	-0.036	-0.057	19.286	0.002
		6	0.001	0.025	19.287	0.004
		7	0.021	0.007	19.553	0.007
		8	-0.057	-0.066	21.427	0.006
		9	0.019	0.050	21.645	0.010
		10	0.042	0.019	22.658	0.012
		11	-0.112	-0.123	30.113	0.002
		12	-0.459	-0.430	154.28	0.000
		13	-0.098	-0.002	159.94	0.000
		14	0.035	0.006	160.68	0.000
		15	-0.039	-0.011	161.61	0.000
		16	-0.061	0.004	163.82	0.000
		17	0.045	0.028	165.01	0.000
		18	-0.009	-0.022	165.06	0.000
		19	-0.004	0.018	165.07	0.000
		20	0.045	0.013	166.29	0.000
		21	-0.026	-0.011	166.68	0.000
		22	-0.036	-0.002	167.44	0.000
		23	0.047	-0.036	168.77	0.000
		24	-0.012	-0.290	168.86	0.000
		25	0.053	0.070	170.54	0.000

图 7 月度气温 12 步差分后的自相关和偏自相关

Figure 7 The ACF and PACF of month temperature after 12 steps difference

在模型辨识时，Akaike[1,4]的信息准则（Akaike Information Criteion）是适用性较广泛的准则，即 AIC 准则，可用来确定最优模型。AIC 准则是拟合精度和参数个数的加权函数，定义如下：

AIC=－2ln（模型的极大似然函数值）+2（模型中未知参数个数）=$-2L(\hat{\beta})+2k$

式中：k 为模型中未知参数个数，$\hat{\beta}$ 为参数的极大似然估计值，$L(\cdot)$为似然函数。

AIC 准则同时体现了残差不相关原则和模型的简介原则相结合，且能够排除建模者的主观因素。

表 1　模型的样本决定系数

Table 1　The adjusted R-squared of model

模型 Model	Adjusted R-squared
$(1,0,0)\times(1,1,0)_{12}$	0.827 10
$(0,0,1)\times(0,1,1)_{12}$	0.603 70
$(1,0,1)\times(1,1,1)_{12}$	0.248 85

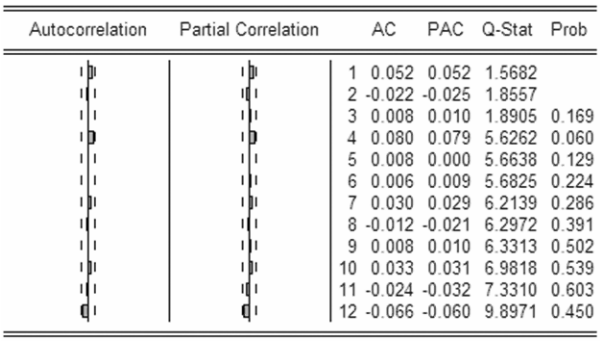


图 8　残差自相关检验

Figure 8　The residual test of ACF

从 D（qw，0，12）的时序图 5 可以看出，其均值稳定，没有明显的周期性；再通过 D（qw，0，12）的相关性分析，如图 7。在 12 步差分后序列自相关和偏自相关图中，序列经过一阶季节差分后，季节性基本消除，故 D=1，尝试使用乘积季节模型 ARIMA$(p,d,q)\times(\mathrm{P,D,Q})_{\mathrm{S}}$ 。不同阶数下拟合模型的样本决定系数值如下表 1。

根据平稳化后数据的自相关图和偏自相关图以及 AIC 准则，建立季节乘积模型，使用最小方差估计法，确定模型口径为：

$$(1-B^{12})x_t=(1+0.080.38B-0.878\,046B^{12})\varepsilon_t$$

2.3　模型检验及预测

模型是否合适，需要对其拟合优度进行检验，典型方法是对观测值和模型拟合值的残差序列进行白噪声检验分析。如果残差序列不是白噪声序列，则说明还有信息包含在相关的残差序列中未被提取，模型其他参数不能完全代表建模对象的统计性质，即所建模型不是最终模型。此时可对残差拟合更复杂的模型，以充分提炼资料的信息，从而得到更合适的模型。ARIMA 季节模型的诊断检验，即模型的残差序列 ε_t 的独立性检验，若通不过检验，需要对原模型进一步改进，再重复以上步骤即可。由图 8 可以看出，ACF 和 PACF 都没有显著等于零，Q 统计量的 P 值都远远大于 0.05。因此，可以认为残差序列为白噪声序列，整个模型提取比较充分。

用此模型对月平均气温数据序列进行拟合并对 2008 年 1 月至 12 月的月平均气温进行预测，如表 2。预测结果显示，预测值与实际值基本吻合，表明模型的选择是正确的，平均绝对误差为 0.875 ，拟合效果较好。

表 2　2008 年预测值

Table 2　The forecasting value of month mean temperature in 2008　　℃

月份 Month	实测值 Measured value	SARIMA 模型 Model	绝对误差 Absolute error	RBF 预测值 Predicted value	绝对误差 Absolute error
1	1.8	2.7	0.9	3.67	1.87
2	3.3	4.5	1.2	5.69	2.39
3	12.8	10.7	2.1	9.95	2.85
4	17.2	17.2	0.0	16.21	1.01
5	23.6	22.0	1.6	21.48	2.12
6	24.4	25.5	1.1	25.05	0.65
7	29.4	29.2	0.2	28.60	0.80
8	27.7	27.9	0.2	28.00	0.30
9	24.7	23.6	1.1	23.36	1.34
10	19.7	18.5	1.2	17.95	1.75
11	12.2	12.4	0.2	11.87	0.33
12	7.0	6.3	0.7	6.00	1.00

3　结论

铜陵地区的月平均气温序列是一非纯随机性时间序列，相关性分析显示其具有明显的以年为周期的周期性。作者对铜陵地区月平均气温的动态数据进行了建模与预报，并和径向基(RBF)神经网络模型的预测值进行了对比，拟合及预报效果较好。但是影响一个地区的气温的因素很多，诸如地区纬度的差异、降水的变化、生态环境的因素（植被覆盖）等。可以尝试将影响气温的诸多因素作为输入变量，

温度作为输出变量进行多变量的时序分析，以便拟合更优的模型。

参考文献：

[1] Geoge E P Box. 时间序列分析预测与控制[M]. 顾岚, 主译. 范金城, 译. 北京: 中国统计出版社, 1997: 101-135.

[2] 王艳. 应用时间序列分析[M]. 北京: 中国人民大学出版社, 2005.

[3] 李鹏飞, 周晓飞, 张庆国, 等. 铜陵市近 49 年气温变化特征及其趋势分析[J]. 安徽农业大学学报, 2010, 37(2): 346-351.

[4] 徐丽, 张庆国. 种群数量动态时序建模方法的研究[J]. 安徽农业大学学报, 1995, 22(3): 320-326.

[5] 田铮. 时间序列分析的理论与应用[M]. 北京: 高等教育出版社, 2003: 214-246.

[6] 薛冬梅. ARIMA 模型及其在时间序列分析中的应用[J]. 吉林化工学院学报, 2010, 27(3): 80-83.

[7] 刘贤赵, 邵金花. 烟台地区降水量的 ARIMA 随机模型研究[J]. 数学的实践与认识, 2006, 36(8): 8-12.

[8] 彭志行, 鲍昌俊, 赵杨, 等. ARIMA 乘积季节模型及其在传染病发病预测中的应用[J]. 数理统计与管理, 2008, 27(2): 362-368.

[9] 郎喜白, 曾黄锦. 湛江地区年降水量的时间序列分析[J]. 水利科技与经济, 2006, 12(8): 505-507.

[10] Chen Y H, Li C L. Parameter estimation and forecast in fractional order ARIMA models [J]. Systems Engineering, 2004, 22(6): 87-90.

[11] 范金城, 梅长林. 数据分析[M]. 北京: 科学出版社, 2002.

[12] 罗凤娟, 董晓萌, 郭满才, 等. 季节性模型在杨凌示范区气温预报中的作用[J]. 中国农学通报, 2007, 23(11): 388-391.

[13] 孔朝莉, 刘双, 杨启昌. 沈阳地区月平均降雨量的 ARIMA 时序建模与预测[J]. 鞍山师范学院学报, 2003, 5(6): 32-34.

本刊外聘编委　岳永德教授

1978－1981 就读于浙江农业大学，获农药残留与环境毒理方向硕士学位；1985 年 11 月－1988 年 2 月在德国 Fraunhofer 环境化学和生态毒理学研究所及霍恩海姆大学植物医学系，访问学者，从事农药残留及环境毒理研究；1993 年 3－10 月在德国萨尔兰大学生物化学和药物化学系，高级访问学者，从事农药代谢降解研究。1994-2003 年任安徽农业大学副校长，2003 年 8 月调任国家林业局国际竹藤网络中心常务副主任。

兼任国家科技部“食品安全重大科技专项”专家组成员，国家竹藤标准化委员会副主任委员，“茶叶生物技术”国家重点开放实验室学术委员会副主任，享受国务院特殊专家津贴。

曾主持完成多项国家和省部级农药残留分析、典型环境污染物转归和农药安全使用标准研究课题。包括国家自然科学基金项目 3 项，国家攀登计划子课题 1 项，安徽省“九·五”、“十·五”攻关项目、安徽省自然科学基金等项目多项。近年主持了国家十五重大科技专项“农药残留检测技术”子项目 2 项，国家 863 项目子项目 1 项。主编全国统编教材《农药残留分析》、《环境保护学》以及《茶叶农药残留与控制》、《有害生物综合治理与展望》、《农药残留研究进展》等著作。发表研究论文 100 余篇，其中 SCI 收录 9 篇。曾获安徽省首届青年科技奖，获省部级科技进步二等奖 2 项、自然科学三等奖 1 项和科技进步三等奖 2 项。