**Holt-Winter指数平滑与ARIMA模型在中国水力发电量预测中的应用**

摘 要：作为一种清洁、可再生的能源形式，水力发电在我国能源结构中占据越发重要的地位。本文首先利用因素分解模型处理我国水力发电量时间序列数据，成功提取了它的趋势效应、季节效应和随机效应，再利用Holt-Winter三参数指数平滑模型和ARIMA乘法模型对该序列进行预测，并得出水力发电量受到水电站的建设与使用、主要河流水量水位的季节性变化因素等影响的结论。本研究不仅有助于增进对我国水力发电行业的深入了解，也为未来的水电资源规划和管理提供了有益的参考。

关键词：因素分解模型，Holt-Winters指数平滑，ARIMA模型，水力发电量

**一、引言**

我国经济飞速发展的同时，对能源的需求不断攀升，推动了对清洁能源的广泛关注和应用。在诸多清洁能源形式中，水力发电因其可再生的特性，在我国能源结构中占有举足轻重的地位。本文主要关注我国水力发电量的时间序列变化，以解答水力发电量是否存在明显的趋势性变化，在时间尺度上是否具有明显的周期性变化，这些变化受到哪些因素的影响等问题。

水力发电在我国能源转型中具有重要的战略地位，对其时间序列数据的深入分析具有重要的理论和实践意义。通过揭示水力发电量的趋势性和周期性特征，不仅能为水力发电行业的管理和规划提供科学依据，还有助于完善我国清洁能源体系。此外，研究结果还将为国家能源政策的制定提供实用的统计支持，推动清洁能源的可持续发展。

**二、数据介绍**

本文所使用的数据来自国家统计局网站，涵盖了1986年至2022年每年3月至12月的水力发电量数据。值得注意的是，在数据选择时省略了1月和2月的记录。这主要有两方面的考虑：1月和2月通常是我国水资源的枯水期，此时河流水位较低，水力发电量相对较小；近年数据中缺少对这两个月水力发电量的记录，给时间序列分析带来挑战。综合考虑上述因素，本文仅研究3至12月数据，以便更好地关注水力发电的季节性和长期趋势，提高研究的准确性和可解释性。

**三、研究方法**

**（一）因素分解模型**

因素分解模型基于确定性因素分解的思想，被广泛应用于时间序列分析中。确定性因素分解认为所有的序列波动都可以归纳为趋势（T）、季节（S）、交易日（D）、随机波动（I）四个因素的综合影响。由此写出表达形式：

具体来说，针对不同特征的时间序列可以有加法模型、乘法模型、伪加法模型、对数加法模型等不同设定形式。利用因素分解模型对水力发电量的时间序列数据进行拆解，可以明确长期趋势的演变、季节性变化的规律以及可能存在的随机波动，为预测提供支持。

**（二）Holt-Winters三参数指数平滑模型**

Holt-Winters模型是一种针对具有长期趋势和季节效应的时间序列数据的预测模型。它具有加法和乘法两种形式。以乘法模型为例，假设序列具有如下结构：

其中，为截距，为斜率，为随机波动，为t时刻季节效应。其可等价表示为如下递推形式：

利用其中三个参数的递推公式、、可得出Holt-Winters三参数指数平滑乘法公式向前k期预测值。

**（三）ARIMA乘法模型**

相较于简单的季节加法模型，ARIMA季节乘法模型能够更好地捕捉序列季节效应、长期趋势效应和随机波动之间的复杂交互影响。乘法模型的原理是将序列的短期相关性和季节相关性分别用低阶和以周期步长为单位的模型提取，并将拟合模型设定为它们的乘积。这种模型设定有助于更全面地理解水力发电量的复杂变化，并为未来水力发电量的预测提供科学依据。

**四、研究结果**

**（一）因素分解模型**

为对我国水力发电量时间序列数据进行可视化，考察1986-2022年我国水力发电量的确定性影响因素，绘制时序图1。

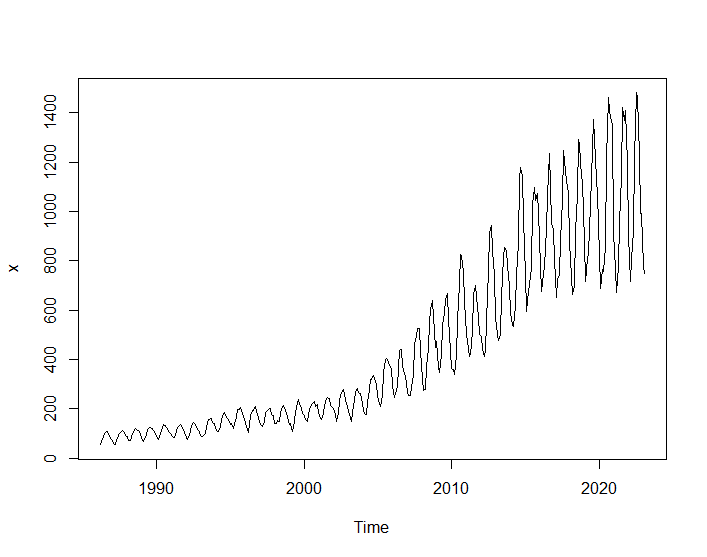


图1 1986-2022年我国水力发电量序列时序图

结果显示，该时间序列具有明显的线性递增趋势以及以10期为周期的季节效应，且没有循环波动或交易日特征，可以初步确定该序列受到长期趋势、季节效应和随机波动三个因素的影响。此外，图1显示出随着趋势的递增，振幅在不断扩大，根据时间序列理论，这反映出季节效应受到趋势的影响。根据以上分析，本文均选用乘法模型来进行拟合和预测。

为将各效应提取和分离，使用R语言中的decompose函数处理数据。将得到的季节指数、趋势效应和随机效应进行可视化，得到图2至图4所示的效果图。

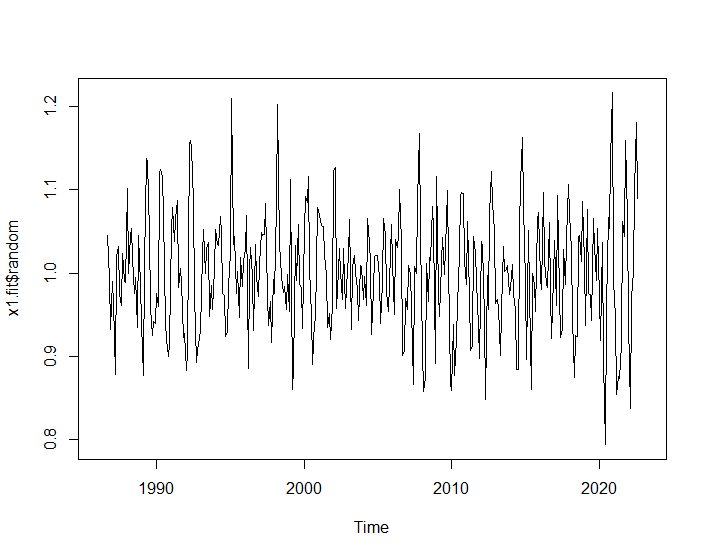
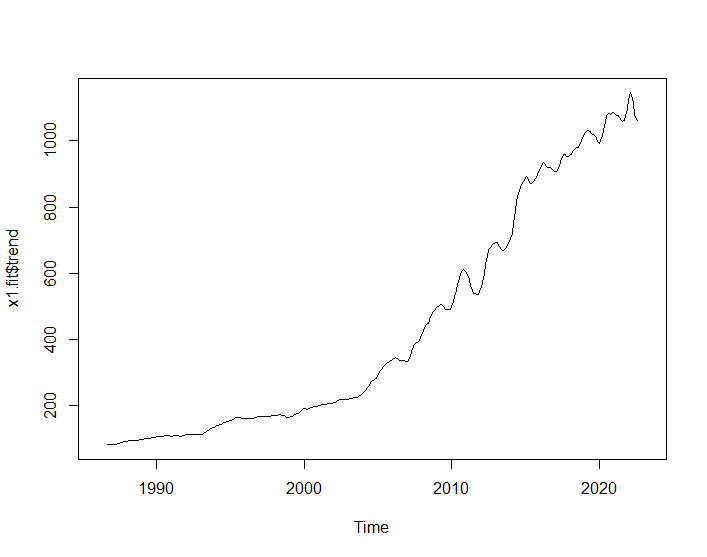
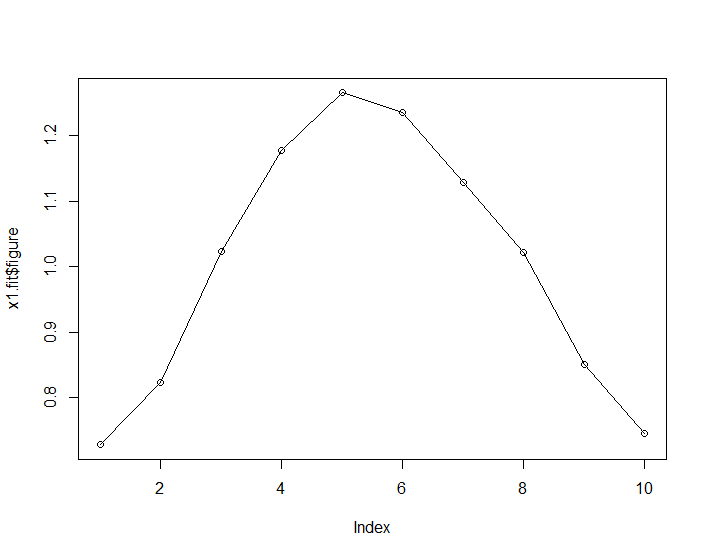


图2 季节指数效果图 图3 趋势效应效果图 图4 随机效应效果图

逐一分析图2至图4的结果，季节指数效果图表明本数据具有明显的季节效应。具体来说，季节指数呈现出夏秋高、冬春低的特征。联系实际，我国水力发电集中在图5所示长江流域的四川、云南、湖北、贵州等地，该地区主要的气候类型为亚热带季风气候，表现为夏季高温多雨，冬季温和少雨。降水量的特征反映在地表径流上，也会呈现出夏季多、冬季少的特点，这与水力发电量的季节指数高度一致。

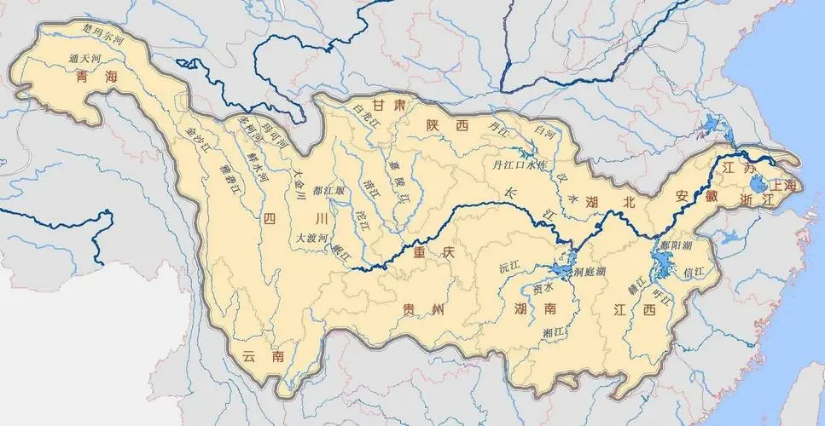


图5 长江流域示意图

自1986年以来，中国水力发电迅速发展，取得显著成就。在这段时间内，水电站的建设速度加快，一系列重要的水电工程相继建成。反映在图3中，我国1986年以来的水力发电量具有明显的增长的趋势效应，这是我国水力发电伟大成就的充分体现。从图4提取出的随机效应可以看出，在将季节指数和趋势效应剔除后，剩余的部分呈现纯随机序列的特点。综合2、3、4图，我国水力发电量时间序列数据中的不同因素能够较好地分解，可以在此基础上选择适当的模型预测序列未来的发展。

**（二）指数平滑预测模型**

由前文分析可知，我国水力发电量数据具有明显的季节效应，结合确定性因素分解理论，选择Holt-Winters三参数指数平滑模型进行序列预测。在不特别指定平滑系数的值的基础上利用HoltWinters函数处理数据，得到基于最优拟合原则的平滑系数取值：

进一步得出三个参数的递推公式，其中c(t)对应的10期季节指数值见表1：

表1 3-12月季节指数

|  |  |  |  |
| --- | --- | --- | --- |
| 月份j | 季节指数 | 月份j | 季节指数 |
| 3 | 0.7444787 | 8 | 1.2227264 |
| 4 | 0.7979603 | 9 | 1.1182969 |
| 5 | 1.0006368 | 10 | 1.0228130 |
| 6 | 1.2132179 | 11 | 0.8100863 |
| 7 | 1.3318692 | 12 | 0.7266412 |

综上，可得该序列向前k期的预测值：

**（三）ARIMA乘法模型**

根据前文对发电量时序图的分析，本序列的季节效应、长期趋势效应和随机波动之间存在较为复杂的交互影响关系，选择用季节乘法模型来更加充分地提取其中的相关关系。根据时序图的特点，尝试分别用1阶差分提取季节和序列本身的相关性。绘制差分后的时序图6。

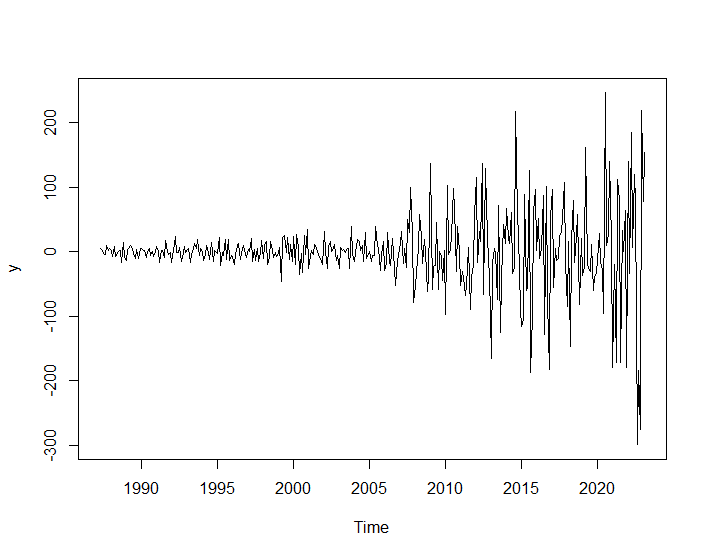


图6 1986-2022年我国水力发电量差分后序列时序图

差分后的时序图已经表现出良好的平稳性特点，为进一步验证，对差分后序列做ADF检验。考虑到原序列的趋势性，采用带常数项和趋势的ADF检验形式。检验结果如表2所示。结果显示均有p <0.05，因此可以认为差分后的序列平稳。为避免差分后序列是白噪声序列，进一步对其进行纯随机性检验。检验结果如表3所示。结果显示均有p <0.05，因此可以认为差分后的序列不是白噪声。

表2 ADF检验结果 表3 纯随机性检验结果

|  |  |  |
| --- | --- | --- |
| 延迟阶数 | 值 | p值 |
| 0 | -19.6 | 0.01 |
| 1 | -13.6 | 0.01 |
| 2 | -12.1 | 0.01 |

|  |  |  |
| --- | --- | --- |
| LB值 | 延迟阶数 | p值 |
| 19.142 | 6 | 0.00393 |
| 96.101 | 12 | 3.22e-15 |
| 108.42 | 18 | 6.217e-15 |

为进行模型识别和定阶，绘制图7所示ACF和PACF图。先分析季节效应，观察图像发现，季节效应的ACF图具有截尾性，而PACF图表现为拖尾，综上对于季节相关性可以使用模型提取。再分析短期相关性，ACF和PACF图均呈现拖尾特征，经尝试使用模型提取较为合适。

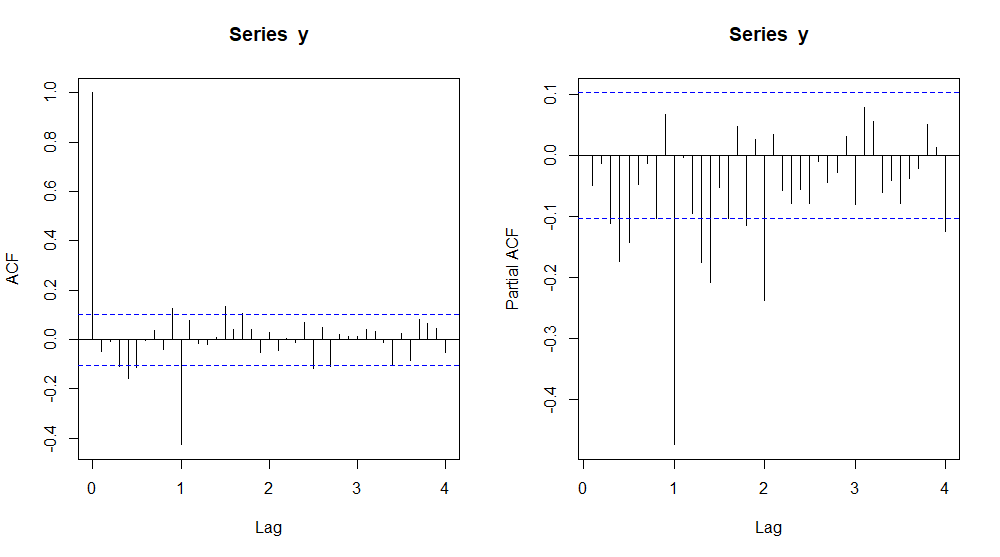


图7 差分后序列自相关和偏自相关图

由于短期相关性和季节效应之间具有乘积关系，所以拟合模型实际上为和的乘积。根据假设对原始数据使用条件最小二乘与极大似然混合估计方法拟合季节乘法模型，得到该模型的拟合口径如下：

对拟合模型进行显著性检验，检验结果如图8。结果显示，模型的残差为白噪声序列，系数均显著非零，这说明该模型拟合良好，对序列信息进行了充分的提取。

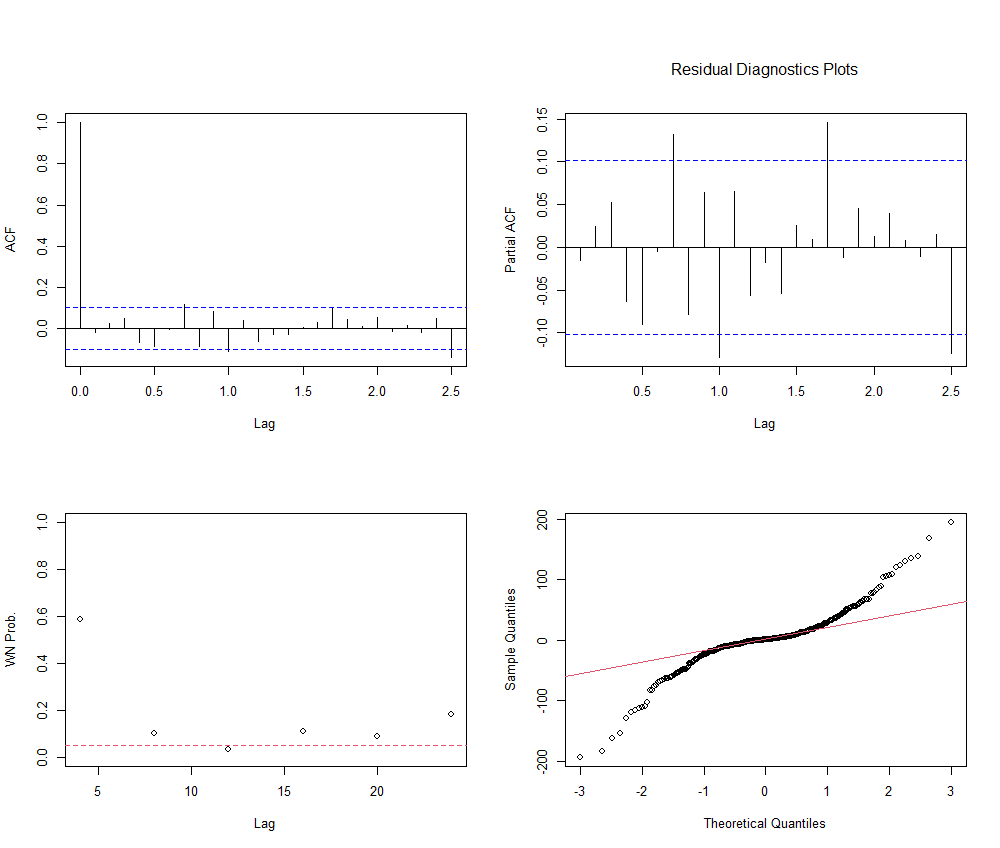


图8 模型显著性检验图

**（四）序列预测**

上述指数平滑预测模型和ARIMA乘法模型均对我国水力发电量数据进行了很好的描述，为了直观呈现两模型的优异的拟合效果，分别将序列拟合值和序列观察值联合作图，如图9、10所示。观察图像可知，真实值和通过模型得到的预测值曲线重合度较高，说明指数平滑预测模型和ARIMA乘法模型的拟合效果均十分良好。另外，由于国家统计局已经公布了2023年的部分数据，可以用于和模型预测结果进行对比，在两幅图中均绘制出了对后5期的预测情况。

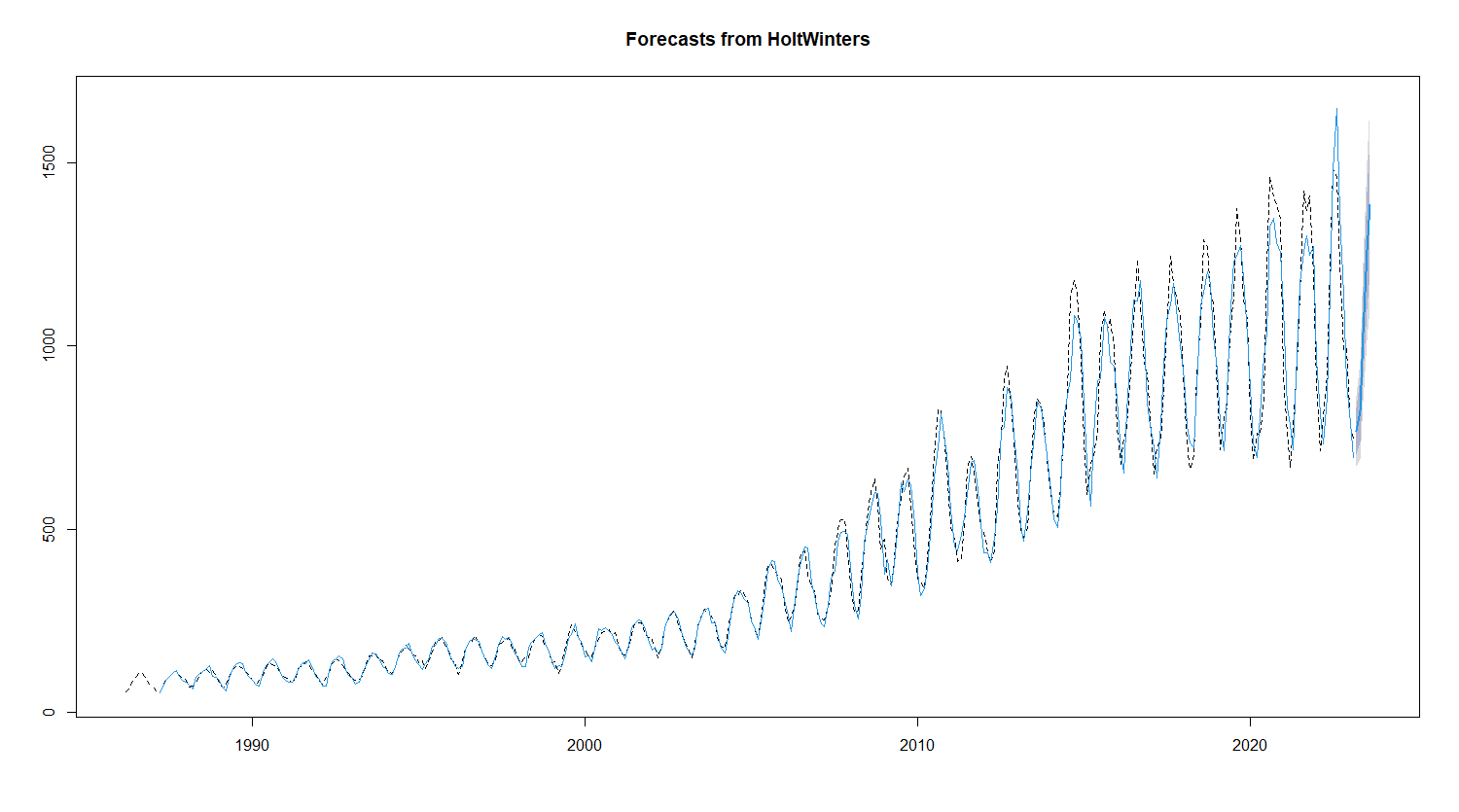


图9 Holt-Winters模型下我国水力发电量序列预测效果图

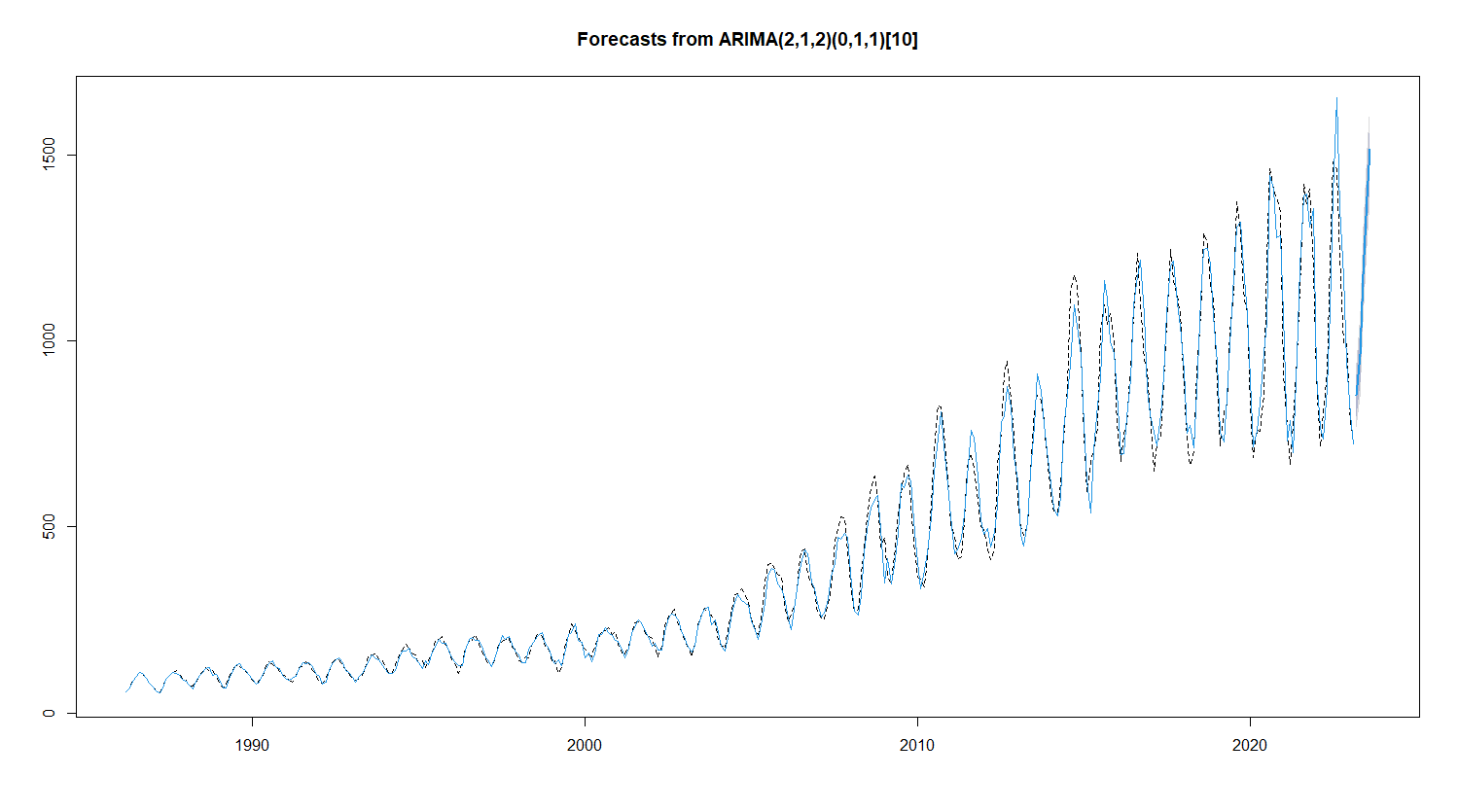


图10 ARIMA模型下我国水力发电量序列预测效果图

为了更好地比较，将真实值和两种预测值在表4中放在一起比较。表中结果表明真实数据和用模型得到的预测数据存在一定的差距。从模型的角度来看，指数平滑模型在面对复杂的非线性关系、时间序列中的突发变化或异常值时，其灵活性可能受到限制，而ARIMA对于数据的平稳性要求相对较高，且需要进行较多的参数调优工作，这些都可能是造成上述差异的原因。另外，水力发电量的实际情况可能受到多种因素的影响，环境因素、政策变化、自然灾害等外部因素可能在一定程度上超出了模型的预测范围。

表4 两模型预测效果比较

|  |  |  |  |
| --- | --- | --- | --- |
| 时间 | 真实数据 | 指数平滑 | ARIMA |
| 2023年3月 | 683.52 | 767.3750 | 853.5394 |
| 2023年4月 | 683.6 | 824.5758 | 964.6636 |
| 2023年5月 | 820.2 | 1036.6137 | 1179.3912 |
| 2023年6月 | 982.12 | 1259.9918 | 1387.6044 |
| 2023年7月 | 1211.3 | 1386.6798 | 1515.0121 |

为进一步评估两模型的预测准确性，引入MAE（平均绝对误差）、MSE（平均平方误差）和RMSE（均方根误差）三个指标，其中MAE表示预测值和实际值之间的平均绝对差值，MSE是对预测值和实际值之间差值的平方，RMSE是MSE的平方根，它们的值越小，说明模型性能越好。

表5 两模型预测效果指标

|  |  |  |  |
| --- | --- | --- | --- |
| 模型 | MAE | MSE | RMSE |
| Holt-Winters | 971.39 | 992415.53 | 996.20 |
| ARIMA | 1096.38 | 1254497.04 | 1120.04 |

表5呈现了两模型的三个指标的值，发现Holt-Winters模型的三个值均较小，因此在本文选用的模型中，Holt-Winters的预测效果更好，这与表4中的结论一致。ARIMA模型表现并不突出，再结合图8对ARIMA模型的显著性检验，虽然检验能够通过，但是并不是非常显著，这表明本文建立的模型中仍有较多信息未能从残差中提取出来，有待进一步优化。

**（五）模型优化**

为充分发挥ARIMA模型在时间序列预测上的能力，解决上文提到的模型有待优化的问题，简要探讨一些优化的方法和策略。首先，本文短期相关和季节相关的差分阶数以及差分后ARMA模型的定阶通过对图像的主观判断得出，在此尝试改用auto\_model函数根据AIC值自动选择最佳模型。函数建议采用来建立模型，结果显示MAE、MSE和RMSE分别为1066.57、1188113.67和1090.01，相比的结果得到一定的优化。另外，可以使用STL方法，使用STL进行季节分解后对序列进行去趋势处理，最后对去趋势序列拟合ARIMA模型，能够得到相同的结果。以上方法的优势在于能够充分发挥计算机的强大运算能力，解决主观判断中遇到的定阶困难等问题，提高模型选择的效率。

**五、研究结论及建议**

本文通过因素分解模型对我国水力发电量的时间序列数据进行了详细分析，成功提取了趋势效应、季节效应和随机效应，通过Holt-Winters三参数指数平滑模型和ARIMA乘法模型的预测分析，证实了这些模型在水力发电量预测方面的有效性。

在进一步的研究中，可以通过残差分析、交叉验证和参数调优等方法来深入探讨模型选择的合理性。在选择最适合实际问题的模型时，需要权衡模型的解释性和预测性能，并结合实际情况，如水电站的具体运行状况、河流水位的实时变化等，进行灵活调整和综合考量。

通过深入研究水力发电量的时间序列变化，可以更好地为水电资源规划和管理提供科学依据，促进我国清洁能源的使用和可持续发展。