科学技术与教育 2026年第3期
ISSN: 3079-4455

# 基于 ARIMA 的醋醅温度短期预测

李丹[1]，李锦松[1,*]，邓岚[2]，韩强[1]
[1]泸州老窖集团有限责任公司，四川泸州，中国
[2]四川轻化工大学自动化与信息工程学院，四川宜宾，中国
*通讯作者

【摘要】在食醋固态发酵工艺中，食醋的产品品质和生产效率会被微生物活性和代谢动态直接影响，其中温度参数的调控是维持微生物代谢稳态的决定性因素。因此如何有效地预测到醋培温度的变化，进而有效地干预就成了当下亟待解决的问题，本文采取基于时间序列的预测模型来实现醋培温度预测。实验选取 2024 年 10 月 16 到 11 月 18 日（共计 34 天）的醋培温度数据，随机选取一口窖池每分钟采样的温度时序数据（总计 47124 个观测点），建立 ARIMA（自回归差分移动平均）预测模型对历史数据进行拟合，实现对醋培温度的短期预测。通过 MAE（平方绝对误差）和 RMSE（均方根误差）来评估模型预测性能。试验结果表明：该模型在一定程度上可以捕捉到温度变化的周期性及变化趋势，其预测值与实际监测值的动态吻合度达 0.8602（Pearson 相关系数），为发酵过程的优化调控提供了有效信息。

【关键词】醋培温度；时间序列的预测模型；ARIMA；卡尔曼滤波；周期性及变化趋势
【基金项目】泸州市科技计划项目（编号：2023SYF157）

## 1.引言

四川麸醋以添加醋药醋曲为主要特色，经过“前稀后固”发酵而成，与山西老陈醋、镇江香醋、永春老醋并列为我国“四大名醋”。[1]经研究表明，影响醋醅中乙酸和乳酸的主要发酵条件是温度和溶解氧，其中乳酸在发酵中前期（1～5d）具有较高的生成速率，乙酸则于发酵中后期（5～9d）快速积累。[2]在实际操作中，由于不同窖池发酵情况差异、常会控温和翻醅时机不精准，导致发酵不完全。[3]因此对醋醅发酵温度的精准预测，可以避免醋醅发酵不完全导致的经济损失，对提高陈醋的合格率具有重要的意义。

醋醅发酵过程中的温度数据呈现出时间序列的主要特征，这些特征具有非平稳性。在时间序列数据的预测领域，现有方法主要可以分为传统预测方法[4,5]和基于人工智能的预测方法[6-12]。人工智能领域主要采用人工神经网络、混沌理论和支持向量机等方法来进行预测和分析。这些技术能够处理复杂的非线性关系，从而提升预测性能。传统方法应用基于统计理论或 ARIMA 模型等方法进行预测；基于神经网络的方法则包括使用人工神经网络、混沌理论和支持向量机等技术。兰永青[6]等提出了一种基于 SSA-LSTM 相结合的预测模型研究了瓦斯浓度预测，实验结果表明该预测有良好的预测精度。王孝东[7]等基于改进蜣螂优化算法优化的 VMD-BiLSTM 模型来预测露天矿坑汛期涌水量，在精度上取得成功。燕学博[8]等通过在 CNN 和 LSTM 的组合模型中 Attention Mechanism（引入注意力机制）去进行货物量的预测，具有较好的预测准确性。杨信廷[9]等基于自注意力机制和独立预测方法构建了 PatchTST 模型，通过融合其与 TiDE 模型中的模块，有效改进了模型预测精度。王静[10]等利用快速傅里叶变换的 Informer 时空预测方法，在时空嵌入层和频域特征提高预测精度。安俊秀[11]针对具有非线性特性的时间序列进行预测，引入可学习的归一化线性变换矩阵，具有很好的理论和应用价值。刘冬兰[12]等针对长时间序列预测不准确、非线性的特点提出了一种基于分解式 Transformer 的联邦长期时间序列预测算法，可以在长序列上准确的预测。

醋醅发酵温度的精准预测不仅可以让作业人员时刻掌握醋醅发酵的真实状态，也能为翻醅机提供精确的翻醅时点，还可以为后续的装坛操作提供方案，对提升陈醋产量具有重要的应用价值。监测数据表明，发酵期间醋醅温度曲线总体呈现出“前期缓慢上升、中期显著升高、后期逐渐下降”的变化趋势，即相邻温度间具有显著的长时间依赖关系。综上所述，面向这种具有非线性趋势的时序

数据，本研究基于自回归差分移动平均模型（ARIMA）建立了一种酒醅温度时间序列预测模型，能够有效利用历史数据来预测未来数据。

## 2.理论介绍

### 2.1 ARIMA 模型

ARIMA 模型是一种经典的时间序列预测模型，主要由自回归（AR）、差分（I）、移动平均（MA）三部分组成。AR 模型描述的是当前值与其过去值之间的关系，在基于时间依赖性和时序衰减这两个重要的假设下，将时间点之间的关系建模为：一个时间点的标签值可以由过来某个时间段的全部标签值的线性组合来表示。其公式为

$$Y_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \xi_t \quad (1)$$

其中：$Y_t$表示在当前时间点 t 的时间序列值，也就是在时间 t 时的标签值（训练时使用真实标签，测试时输出预测标签），$X_{t-1}$、$X_{t-2}$、…、$X_{t-p}$同上，$c$ 为常数，$\varphi_1$、$\varphi_2$、…、$\varphi_p$分别表示为$X_{t-1}$、$X_{t-2}$、…、$X_{t-p}$的自回归系数，$\xi_t$为误差项，也称为白噪声项，$p$是自回归模型中的参数，它表示在预测当前时刻值时，将考虑过去多少个时刻的值。MA 模型描述的是当前时间序列与过去噪声之间的联系，定义基于白噪声的假设而来，要求每个时间点的数据是相互独立且满足相同分布，在给定一个白噪声序列$\varepsilon_t$后，其公式为

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (2)$$

其中：$\mu$是时间序列的均值或期望值，这个值对于所有时间点都是相同的，$\varepsilon_t$、$\varepsilon_{t-1}$、…、$\varepsilon_{t-q}$为白噪声项，$\theta_1$、$\theta_2$、…、$\theta_q$分别表示为不同白噪声项的移动平均系数，$q$是移动平均模型中的参数，表示过去有多少的白噪声项被纳入模型。

ARIMA（自回归差分移动平均）模型可以通过数据的自相关性和差分的形式，提取出隐藏在数据背后的时间序列模型，并用来预测未来的数据。在结合 AR 模型与 MA 模型后，ARIMA 不但可以捕捉到数据的趋势变化，又可以处理那些临时突发的变化或噪声较大的数据。

### 2.2 卡尔曼滤波

卡尔曼滤波主要用于估计动态系统的状态，通过预测和更新两个状态来不断修正状态估计。原始数据在测量时，由于传感器测量问题，导致部分数据缺失。预测过程主要是利用标准更新步骤建立对当前状态的先验估计，及时向前推算当前状态变量和误差协方差估计的值，以便为下一个时间状态构造先验估计值；更新过程负责反馈，利用测量更新方程在预测过程的先验估计值及当前测量变量的基础上建立起对当前状态的改进的后验估计[13]。本次研究采用卡尔曼滤波模型去处理缺失时序数据，可以有效填补数据空缺并维持估计的连续性。

原理主要分为以下三部分：

（1）当数据缺失时，滤波器会依赖状态转移模型来进行预测，状态预测方程为

$$x_{k|k-1} = F_k x_{k-1|k-1} \quad (3)$$

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \quad (4)$$

其中：$x_k$为系统状态，$F_k$为状态转移矩阵，$Q_k$为过程噪声协方差，$P_{k|k-1}$为预测协方差矩阵，表达了当前状态会因过程噪声和模型误差的累积而受到影响。

（2）当时间 k 的观测值$Z_k$缺失，滤波器会跳过卡尔曼增益计算与状态修正步骤，直接保留预测结果为当前最优估计。

$$x_{k|k} = x_{k|k-1} \quad (5)$$

$$P_{k|k} = P_{k|k-1} \quad (6)$$

（3）当后续时刻重新获得有效观测值$Z_{k+m}$时，滤波器通过标准更新步骤融合新观测信息。

$$K_{k+m} = P_{k+m|k+m-1} H_{k+m}^T \left( H_{k+m} P_{k+m|k+m-1} H_{k+m}^T + R_{k+m} \right)^{-1} \quad (7)$$

$$x_{k+m|k+m} = x_{k+m|k+m-1} + K_{k+m} \left( Z_{k+m} - H_{k+m} X_{k+m|k+m-1} \right) \quad (8)$$

其中：$H_{k+m}$为观测矩阵，$R_{k+m}$为观测噪声协方差，卡尔曼增益$K_{k+m}$根据积的不确定性（$P_{k+m|k+m-1}$）动态调整模型预测与观测的权重，实现长期缺失后状态估计的快速收敛。

### 2.3 评价指标

采用平方绝对误差（Mean Absolute Error，MAE）和均方根误差（Root Mean Squared Error，RMSE）来作为温度预测的性能度量指标，它们可以反映出与实际值的差距，也能反映出测量值的可信程度。

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \quad (9)$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (10)$$

**3.模型构建方法**

在 ARIMA 模型中，自回归部分描述的是当前变量与其过去值之间的关系；差分部分可以改变原时间序列的平稳性；移动平均部分描述的是当前时间序列与过去噪声之间的联系，利用滞后残差调整预测值。可以由自回归阶数 $p$、差分次数 $d$、移动平均阶数 $q$ 这三个参数的整合建立起 ARIMA 模型。

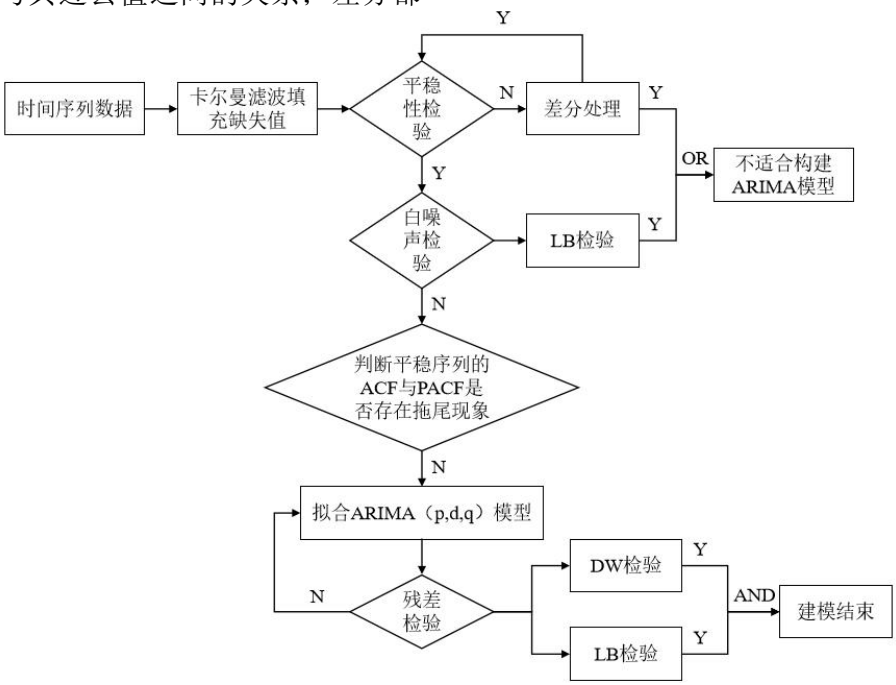


**图 1.构建 ARIMA 模型的流程图**

如图 1 的流程图所示，建立 ARIMA 模型首先对时间序列数据进行预处理，主要是对缺失值进行卡尔曼滤波处理。然后对数据进行平稳性检测，若非平稳性序列，则进行 $d$ 阶差分平稳性处理并检验平稳后的时间序列是否为白噪声序列，若非白噪声，则利用自相关函数（Autocorrelation Function，ACF）图与部分自相关函数（Partial Autocorrelation Function，PACF）图的截尾特性来确定 $p$ 与 $q$ 值，并结合赤池信息准测（Akaike Information Criterion，AIC）来热力图来进一步优化。最终若通过残差检验，即可确定 ARIMA（$p$,$d$,$q$）模型。

**4.数据来源及处理方法**

**4.1 醋培温度数据来源**

实验数据采集于 2024 年 10 月 16 到 11 月 18 日间某市食醋酿造厂生产区域内，该厂里有 16 口醋酸发酵池，随机选取了 1 口酵池测量温度。在这个时间段内醋工艺流程进入到醋酸发酵阶段，由于在过杓和露底阶段翻醅机造成了采集温度的难度，所以利用八通道无纸记录仪和 8 个 PT100 铂电阻温度传感器来进行数据采集，由于露底过程中热量是从上向下传递的，所以温度传感器分别放置在深度 15cm、30cm、45cm、60cm 处，每个温度探测器表面进行抗酸化处理，进而保证了灵敏度和。每个温度传感器所采集的醋醅温度可通过数据线传输至无纸记录仪，温度模块将模拟信号转换为数字信号并通过 RS232 传输到无纸记录仪中。醋醅的温度每分钟记录 1 次，在醋醅发酵过程中的 31 天中，每天 24 小时不间断记录。表 1 为 8 个通道的温度可视化数据。

**表 1.通道的部分温度数据**

| 序号 | Timestamp | channel |
|---|---|---|
| 1 | 2024-10-16 17:25 | 25.3 |
| 2 | 2024-10-16 17:26 | 25.3 |
| 3 | 2024-10-16 17:27 | 25.3 |
| 4 | 2024-10-16 17:28 | 25.3 |

**4.2 数据预处理**

由于传感器在测量期间曾短暂出现问题，导致某些时间段缺失数据，如图 2 图 3 所示，而时序预测对时间序列数据的连续性有着不可或缺的依赖性。为提升模型训练与预测的准确性，采用卡尔曼滤波对缺失数据进行补充，有效填补数据空缺并维持估计的连续性，并保证补充后的数据的时间戳的一致性。

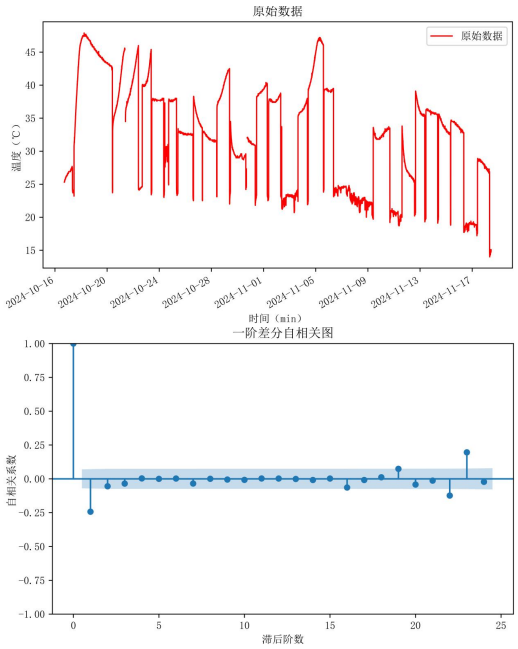


**图 2.醋培温度数据预处理前的结果（采集频率 1min）**

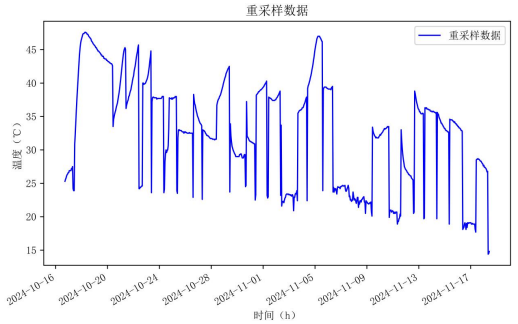


**图 3.醋培温度数据预处理后的结果（采集频率 1h）**

### 4.3 划分数据集

将数据集按时间顺序划分为训练集与测试集。训练集用来训练模型，测试集用于验证真实值与预估值之间的差异，并评估模型的性能与准确性。短期预测旨在预测未来九天的温度值，将 47123 条数据后 30%的数据作为预测目标，用于评估模型的泛化能力。

## 5.基于 ARIMA 模型的短期预测

ARIMA 模型对时间序列数据的平稳性有着严格的要求，为了使模型有更好的预测能力，采用 ADF 检验与 KPSS 检验来评估原始数据的稳定性。结果如表 2，可得：原始数据的 ADF 检验 $p$ 值 2.265E-3，KPSS 检验 $p$ 值 0.01，不符合平稳性要求，而一阶差分后的数据满足平稳性要求。因此，拟采用一阶差分后的数据来构建 ARIMA 模型，即 ARIMA（$p$，1，$q$）。

在判断平稳性后，对处理后的一阶差分数据进行 Ljung-Box 检验，结果如表 3，12 阶滞后的 $p$ 值均小于 0.05，表明该时间序列非白噪声。原醋培温度数据与一阶差分后数据对比如图 4 所示，故正式采用一阶差分后的数据来构建 ARIMA 模型。

**表 2.醋培温度数据的平稳性检验结果**

| | 原始数据 | 一阶差分数据 |
|---|---|---|
| ADF 检验 p 值 | 2.265E-3 | 6.05E-12 |
| KPSS 检验 p 值 | 0.01 | 0.1 |

**表 3.对一阶差分数据进行的 LB 检验**

| | lb_stat | lb_pvalue |
|---|---|---|
| 1 | 46.457582 | 9.362341e-12 |
| 2 | 48.827708 | 2.495732e-11 |
| 3 | 49.806286 | 8.785278e-11 |
| 4 | 49.813688 | 3.949205e-10 |
| 5 | 49.813703 | 1.512925e-09 |
| 6 | 49.818325 | 5.112242e-09 |
| 7 | 50.780291 | 1.014870e-08 |
| 8 | 50.780450 | 2.892392e-08 |
| 9 | 50.802996 | 7.605843e-08 |
| 10 | 50.850000 | 1.861323e-07 |
| 11 | 50.855117 | 4.390825e-07 |
| 12 | 50.859744 | 9.860945e-07 |

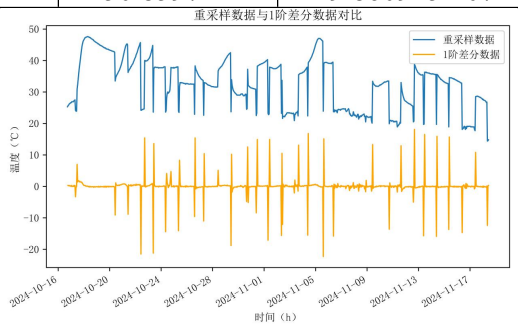


**图 4.原醋培温度数据与一阶差分数据对比**

**图 5.一阶差分数据的 ACF 图**

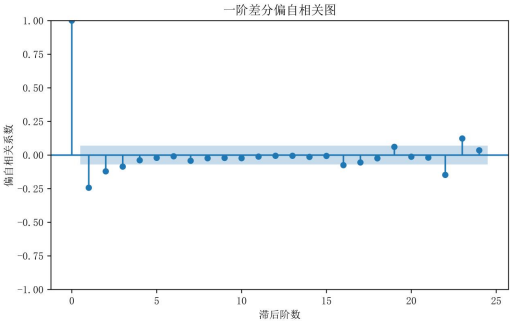


**图 6.一阶差分数据的 PACF 图**

通过构建一阶差分数据的 ACF 与 PACF 图，可以大致得到 ARIMA（$p$，1，$q$）模型中的 $p$ 值与 $q$ 值。结果如图 5 图 6 所示，可以看出：一阶差分数据的 ACF 与 PACF 图均未出现拖尾现象，故应采用参数 $p$、$q$ 均不为

零的 ARIMA 模型来拟合数据。

为了更加准确地得出模型的 $p$ 值与 $q$ 值，分别设置 $p$、$q$ 参数的范围为[0，3]，通过 AIC 热力图来计算所有组合可能的 AIC 值。结果如图 7。

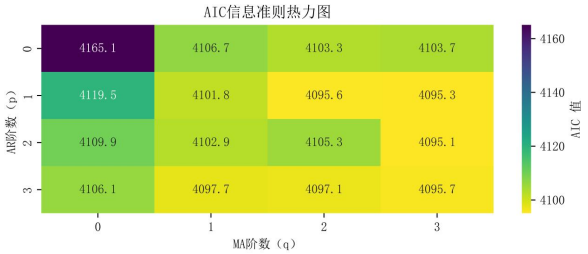


**图 7.醋培温度数据不同 p、q 组合的 AIC 图**

可以看出，当 $p$、$q$ 参数组合为（1，2），（1，3），（2，3），（3，3）时，AIC 值均较小且相差不大，使用上述参数进一步构建 ARIMA 模型。通过比较不同参数下的 MAE 值，结果如表 4，最终选择参数 $p$、$q$ 为（1，3）来作为温度序列预测的 ARIMA 模型。

**表 4.不同（p，q）参数下的 MAE 值**

| | MAE 参数 |
|---|---|
| （1，2） | 1.3293 |
| （1，3） | 1.3239 |
| （2，3） | 1.3808 |
| （3，3） | 1.3679 |

**表 5.对残差序列进行的 LB 检验**

| | lb_stat | lb_pvalue |
|---|---|---|
| 1 | 0.001438 | 0.969751 |
| 2 | 0.017208 | 0.991433 |
| 3 | 0.348432 | 0.950677 |
| 4 | 0.416207 | 0.981129 |
| 5 | 0.485307 | 0.992651 |
| 6 | 0.533219 | 0.997410 |
| 7 | 0.915904 | 0.996073 |
| 8 | 0.928079 | 0.998663 |
| 9 | 0.933149 | 0.999577 |
| 10 | 0.940716 | 0.999870 |
| 11 | 1.053722 | 0.999934 |
| 12 | 1.177462 | 0.999965 |

当得到 $p$、$q$ 参数为（1，3）的 ARIMA 模型后，对其残差序列进行验证，使用 Ljung-Box 检验，结果如表 5，并画出残差图与核密度图，结果如图 8。可得：12 阶滞后的 p-value 值均大于 0.05，残差图大致分布均匀且核密度图基本符合正态分布，表明残差序列无自相关性，符合白噪声特征。因此：可以确定 ARIMA（1，1，3）模型可以有效地捕捉到温度序列变化趋势。

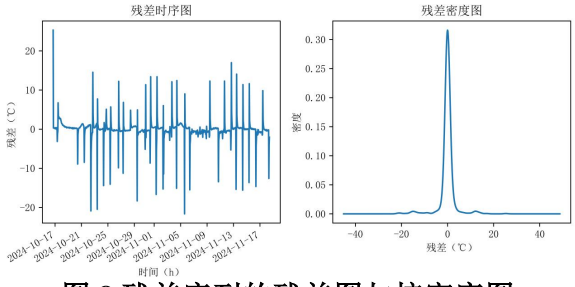


**图 8.残差序列的残差图与核密度图**

基于处理好的温度序列历史数据，拟合 ARIMA（1，1，3）模型，利用该模型对温度序列数据做短期预测，结果如图 9 图 10 所示，其中红线为真实值，蓝线为预测值。结果表明：预测值与实际监测值的 MAE 值为 1.3239，RMSE 值为 3.3189，Pearson 相关系数达 0.8602。

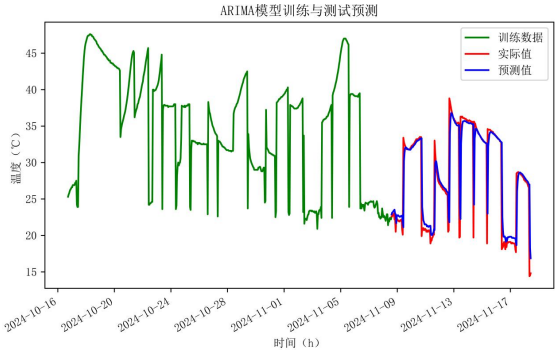


**图 9.ARIMA 模型对温度序列进行短期预测**

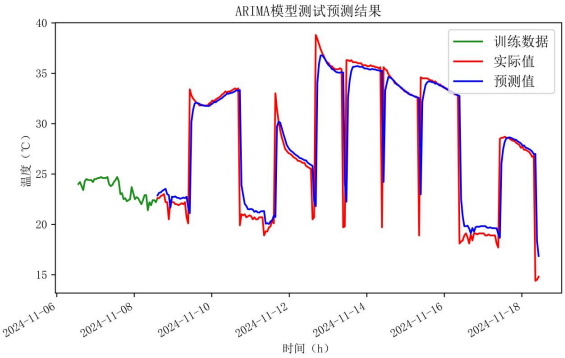


**图 10.ARIMA 模型对温度序列进行短期预测（局部放大图）**

表 6 汇总了 ARIMA（1,1,3）模型的参数估计结果。具体而言，模型的一阶自回归项（ar.L1）系数估计值为 0.9497，前两阶移动平均项（ma.L1,ma.L2）系数分别为-1.2540 和 0.2024。统计分析表明，上述项的 P 值均小于 0.001，在统计学上高度显著，且其 95%置信区间均未包含零值，证明这些滞后变量对序列具有显著的解释力。相比之下，第三阶移动平均项（ma.L3）系数为 0.0558，其 P 值为 0.080，未达到显著性水平，表明该高阶项对模型拟合精度的提升作用有限。此外，

模型残差方差（sigma2）估计值为 10.6439，具有统计显著性。总体而言，该 ARIMA（1,1,3）模型参数估计结果稳健，主要关键系数均呈现出高度的统计显著性，模型能够有效捕捉序列的动态演化特征，适合用于后续的分析与预测。

**表 6. ARIMA（1，1，3）模型拟合效果摘要**

| | coef | std err | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|
| ar.L1 | 0.9497 | 0.024 | 0.000 | 0.902 | 0.997 |
| ma.L1 | -1.2540 | 0.025 | 0.000 | -1.303 | -1.205 |
| ma.L2 | 0.2024 | 0.038 | 0.000 | 0.128 | 0.277 |
| ma.L3 | 0.0558 | 0.032 | 0.080 | -0.007 | 0.118 |
| sigma2 | 10.6439 | 0.270 | 0.000 | 10.114 | 11.174 |

## 6.结论与建议

由于醋培温度数据与时间序列有着紧密的联系，结合在 10 月到 11 月采集到的醋培温度数据，通过差分法、ACF、PACF 图和 AIC 热力图来确定 $p$、$d$、$q$ 值，使用 ARIMA 时序预测模型来建立醋培温度预测模型，并采用 MAE 和 RMSE 对模型进行评估。结果表明：使用 ARIMA（1，1，3）模型对醋培温度数据进行建模后，MAE 值为 1.3239，RMSE 值为 3.3189，Pearson 相关系数为 0.8602。ARIMA 模型虽是一个基础的时序预测模型，其用在醋培温度的预测却较少。本文通过醋培历史温度数据进行了 ARIMA 模型的建立，从结果可以得到，在短时间的预测中，该模型能大致拟合出温度序列数据的变化趋势。由此可见，ARIMA 模型用在醋培温度监测上是可行的，但由于方法的局限性，预测精度还有待提升。为了达到更好的预测作用，可以采取多变量时序预测，通过提取多个特征值，例如：温度、溶解氧含量等，以此能更好地捕捉到醋培温度序列的趋势，达到更高精度的预测；或将 RNN、LSTM 等机器学习模型与时序预测模型相结合，可以有效地捕捉到时序模式，提升预测性能。

## 参考文献

[1] 李宇薇，帖余，唐之兴，等.四川晒醋特征风味物质鉴定及其陈酿前后变化研究[J/OL].食品与发酵工业，2024，1-11.
[2] 张强，赵翠梅，李晓伟，等.温度和翻醅对食醋固态发酵产酸的影响[J].中国酿造，2020，39（04）：159-164.
[3] 朱瑶迪，邹小波，徐艺伟，等.镇江香醋固态发酵过程中温度的监控与分析[J].中国食品学报，2016，16（02）：124-129.
[4] 李鹏，朱洪泽，骆光杰，等.基于 ARMA 模型的海上风机随机风场模拟[J].武汉大学学报（工学版），2024，57（01）：112-120.
[5] 赵强，王擎宇，舒志光.基于 SARIMA 模型的近岸海表温度短期预报研究[J].海洋预报，2024，41（01）：42-49.
[6] 兰永青，乔元栋，程虹铭，等.基于 SSA-LSTM 的瓦斯浓度预测模型[J].工矿自动化，2024，50（02）：90-97.
[7] 王孝东，杨懿杰，吕玉琪，等.改进多变量时序模型的露天涌水量预测[J].安全与环境学报，2024，24（08）：2994-3004.
[8] 燕学博，曹世鑫.基于 CNN-LSTM-Attention 组合模型对我国货运量时序预测对比[J].物流科技，2024，47（14）：5-9.
[9] 杨信廷，郭向阳，韩佳伟，等.基于 TiDE-PatchTST 模型的柑橘冷藏效率时序预测模型优化[J].农业机械学报，2024，55（07）：396-404.
[10] 王静，刘瑞，杨松涛，等.改进 Informer 模型的苜蓿土壤湿度预测方法[J/OL].计算机技术与发展，2024，1-7.
[11] 安俊秀，万里浪. StabilizeNet：用于缓解时间序列非平稳性的新型框架[J/OL].大数据，2024，1-15.
[12] 刘冬兰，刘新，刘家乐，等.基于分解式 Transformer 的联邦长期时间序列预测算法[J].山东大学学报（工学版），2024，54（05）：101-110.
[13] 彭丁聪.卡尔曼滤波的基本原理及应用[J].软件导刊，2009，8（11）：32-34.