

# 基于ES-GARCH方法的COVID-19疫情预测

马永华, 秦喜文\*, 生 菡, 董小刚

长春工业大学数学与统计学院, 吉林 长春

Email: \*qinxiwen@ccut.edu.cn

收稿日期: 2021年4月25日; 录用日期: 2021年5月8日; 发布日期: 2021年5月28日

## 摘 要

新型冠状病毒COVID-19导致的呼吸系统疾病在全球范围内爆发。为了更好地实施防控疫情的相关决策, 研究人员利用统计模型预测疫情发展趋势以及评价这次传染病所带来的影响。在本文中, 提出自回归条件异方差(ES-GARCH)的误差修正模型首次用来预测COVID-19的死亡人数和确诊病例, 并综合了其他六种模型来探索预测精度最高的模型。首先, 选取湖北省新冠肺炎的累计确诊人数和累计死亡人数, 共两组数据集; 其次, 将数据集分别划分为训练集和测试集来建立模型; 最后, 使用评价指标对所使用模型的预测精度进行评估, 结果显示本文中所提出的ES-GARCH模型的预测结果最佳。此外, 为了确保模型的可靠性, 选用意大利的两个数据集对模型进行验证, 结果表现为该模型的预测性能优于其他六个预测模型。在文章的最后, 使用ES-GARCH模型预测湖北省未来七天内COVID-19的累计确诊病例和累计死亡人数。通过验证, 该模型适用于时间序列的短期预测, 对预测全球疫情发展趋势有重要意义。

## 关键词

COVID-19, ES-GARCH模型, 确诊病例, 累计死亡人数, 评价指标

# Prediction of COVID-19 Epidemic Situation Based on ES-GARCH Method

Yonghua Ma, Xiwen Qin\*, Han Sheng, Xiaogang Dong

School of Mathematics and Statistics, Changchun University of Technology, Changchun Jilin

Email: \*qinxiwen@ccut.edu.cn

Received: Apr. 25<sup>th</sup>, 2021; accepted: May 8<sup>th</sup>, 2021; published: May 28<sup>th</sup>, 2021

\*通讯作者。

文章引用: 马永华, 秦喜文, 生菡, 董小刚. 基于ES-GARCH方法的COVID-19疫情预测[J]. 应用数学进展, 2021, 10(5): 1702-1712. DOI: 10.12677/aam.2021.105181

## Abstract

Respiratory diseases caused by New Coronavirus COVID-19 break out all over the world. In order to better implement the relevant decision-making of epidemic prevention and control, researchers use statistical models to predict the trend of epidemic development and evaluate the impact of the epidemic. In this paper, an error correction model of autoregressive conditional heteroscedasticity (ES-GARCH) is proposed to predict the number of deaths and confirmed cases of COVID-19 for the first time, and other six models are integrated to explore the model with the highest prediction accuracy. First, the cumulative number of people diagnosed with COVID-19 and the cumulative number of deaths in Hubei Province were selected, and there were two data sets; secondly, we divide the data set into training set and test set to build the model; finally, the evaluation index is used to evaluate the prediction accuracy of the model, and the results show that the ES-GARCH model proposed in this paper has the best prediction results. In addition, in order to ensure the reliability of the model, two Italian data sets are selected to verify the model. The results show that the prediction performance of the model is better than the other six prediction models. At the end of the paper, ES-GARCH model is used to predict the cumulative number of confirmed cases and deaths of COVID-19 in Hubei Province in the next seven days. Through validation, the model is suitable for short-term prediction of time series, and is of great significance to predict the development trend of global epidemic.

## Keywords

COVID-19, ES-GARCH Model, Confirmed Case, Cumulative Death Toll, Performance Index

Copyright © 2021 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

在 2019 年 12 月底, 由新型冠状病毒(COVID-19)引起的非典型肺炎得到人们的广泛关注[1]。新型冠状病毒是冠状病毒的一种, 是一类严重影响人类身体健康的疾病, 能够感染人体的多个器官, 比如肝脏、肠胃、神经系统, 严重影响人体的呼吸系统[2]。自发现第一例新型冠状病毒携带者至 2020 年 3 月 28 日, 中国累计确诊人数高达 81439 例, 累计死亡人数高达 3300 例。其中武汉市作为国内首个发现确诊病例的地方, 疫情最为严重, 截至 2020 年 3 月 28 日, 武汉累计确诊病例人数达 67801 例, 累计死亡人数高达 3182 例。在[3]中, 作者提到新型冠状病毒在早期的传播能力与严重急性呼吸系统综合征(SARS)相似, 但是从目前情况来看新冠肺炎的发病率已经远远超过 2002 年到 2003 年报告 SARS 的发病率, 对人体的伤害却非常大, 病情严重的患者甚至会有生命危险。为了阻止病毒传播, 防止更多人被感染, 中国政府迅速展开抗击疫情行动, 自 2020 年 2 月 23 日至 24 日以来, 对武汉以及周边主要城市积极采取措施。在疫情初期, 相关领导人果断做出一系列决策, 比如关闭所有娱乐性场所、火车站以及在机场设立体温检测处以便筛查体温异常、对于确诊病例集中收治、对于疑似病例进行隔离、对于密切接触过确诊病例的人员进行为期 14 天的观察等。由于国内疫情初期大多数确诊病例来自武汉, 政府在 2020 年 1 月 23 日上午 10 时, 做出重要决定——对武汉市采取封城管理, 禁止任何人进出, 同时停止一切活动, 包括娱乐性场所、机场、火车、公交地铁等[4]。由于确诊患者在初期迅速增加, 导致医院床位出现供不应求的情况, 政府迅速采取一系列措施, 在十天左右时间建立起两家医院包括火神山医院和雷神山医院, 分别可容纳

1600 张床位和 1000 张床位[5]，及时减轻了武汉市医疗负担。但是由于 COVID-19 的潜伏期时间可以长达 14 天，仍有被感染的人群尚未发现，导致确诊人数不断增加，因此及时有效地分析病情增长趋势，以及预测未来几天的确诊病例数量以及累计死亡人数，对有针对性地采取措施以及制定必要的保护措施对控制疫情的发展是至关重要的[6]。

Domenico Benvenuto [7]等人使用差分整合移动平均自回归模型来拟合 2020 年 1 月 20 日至 2020 年 2 月 10 日 COVID-2019 日流行数据，目的是预测未来两天 COVID-19 的发病率与流行率，其结论表明未来两天的流行率趋于平缓，发病率出现下降趋势；Zixin Hu [8]等人利用 2020 年 1 月 11 日至 2 月 27 日中国确诊病例建立模型，选择改进的堆叠自动编码器来预测累计确诊的趋势线，由此得出结论 COVID-19 的流行将在 4 月中旬结束；Albertine Weber [9]等人使用 SIR 模型预测中国 COVID-19 的流行趋势，结果表明中国 COVID-19 成亚指数型增长；在[10]中，作者建立了一种 COVID-19 的死亡风险评估模型，他们的出的结论是在两种不同情况下死亡率分别是 5.1%和 8.4%，且两种不同情况下的传染率分别是 2.1%和 3.2%，结果表明 COVID-19 可能导致大规模传播；赵[11]等人通过建立一个数学模型来估计 2020 年 1 月上半月未报告的实际 COVID-19 病例数量，作者的出的结论是 2020 年 1 月 1 日至 15 日，未报告的实际确诊病例数 469 起，同时作者根据模型预测在 2020 年 1 月 17 日之后，确诊病例数将会增加两倍。

短期预测可以评估可能的流行趋势，以便于及时修改现有的防护措施[12]。时间序列分是一种科学的定量预测方法，可以根据已有的时间序列数据预测未来的变化，被广泛应用在各个领域，包括预测医疗费用[13]、预测住院人数[14]、预测石油价格[15]等。此外，在以往的传染病研究中常用季节性自回归综合移动平均模型(SARIMA)来实时监测数据如肺结核[16]、腮腺炎[17]、和梅毒[18]等。Arul Earnest 等人使用 ARIMA 模型预测新加坡某三级医院在 SARS 爆发期间的床位的占用情况，得出结论 ARIMA(1,0,3)可以用来很好地描述和预测 SARS 爆发期间的床位占用情况，可以为医院相关领导指定决策提供可靠依据[19]。相关分析指出指数平滑作为经典的时间序列分析模型，也被广泛应用于传染病研究中[20]。

本文的以下章节内容安排如下：第二章详细介绍了本文预测所选用的方法包括整合移动平均自回归模型(ARIMA)、支持向量回归(SVR)、广义自回归条件异方差模型(GARCH)、指数平滑(ES)等。另外，在本文中用来评价模型性能的三个评价指标也在第二章中有详细介绍。在第三章，我们对所选数据进行实证分析，并对比实验结果。在文章结束的部分，对本文的研究结果进行总结，并提出 ES-GARCH 模型可以有更广泛的应用。

## 2. 方法

### 2.1. 整合移动平均自回归模型(ARIMA)

时间序列是指某一指标按照事件发生的先后顺序记录的一组数据。时间序列预测是一种科学的定量预测方法，是一种根据历史数据预测未来数据的分析方法。

整合移动平均自回归模型(ARIMA)是一种经典的分析时间序列数据的方法，也是一种常见的对时间序列进行预测的模型。常见的时间序列模型还有自回归模型(AP)、移动平均模型(MA)、自回归 - 滑动平均混合模型(ARMA)。

ARIMA 模型具有如下结构：

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s x_t) = 0, \forall s < t \end{cases} \quad (2.1)$$

## 2.2. 支持向量回归(SVR)

支持向量回归(SVR)是支持向量机的一个重要分支,所寻求的是所有样本点到平面的总距离最小。其核心思想是通过非线性变换  $\phi(x)$  将输入向量映射到一个高维空间  $H$  中,然后在高维空间中进行线性回归,公式如下:

$$f(x) = \omega * \phi(x) + b \quad (2.2)$$

其中  $\phi(x)$  为特征向量空键,  $\omega$  为权重系数,  $b$  为截距。SVR 拟合回归问题可以表示为目标函数优化问题,遵循结够风险最小化原则,其目标函数公式(2.3)所示:

$$R(x) = \frac{1}{2} \|\omega\|^2 + \frac{1}{m} \sum_{i=1}^m |f(x_i) - y_i|_\varepsilon \quad (2.3)$$

SVR 的实际上是求解凸二次规划问题,因此可以转化为公式(2.4):

$$R(\omega, \xi_i, \xi_i^*) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*)$$

$$s.t. \begin{cases} y_i - \omega\phi(x) - b \leq \varepsilon + \xi_i \\ \omega\phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, m \end{cases} \quad (2.4)$$

其中  $\xi_i$  和  $\xi_i^*$  是松弛变量,  $C$  是惩罚系数,  $\varepsilon$  是不敏感损失函数。引入拉格朗日乘子,最终得到 SVR 模型,如公式(2.5)所示:

$$f(x) = \sum_{i=1}^m (\alpha_i^* - \alpha_i) K(x_i, x) + b \quad (2.5)$$

其中,  $K(x_i, x) = \phi(x_i) * \phi(x)$  为核函数,  $\alpha_i^*$  与  $\alpha_i$  均为拉格朗日乘子。

## 2.3. 广义自回归条件异方差模型(GARCH)

广义自回归条件异方差模型(GARCH)是由 Bollersle (1986)在 ARCH 上发展起来的,是一个非常经典的时间序列拟合模型。ARCH 模型的实质是使用残差平方序列的  $q$  阶移动平均拟合当期异方差函数值,由于移动平均模型具有自相关系数  $q$  阶截尾性,所以 ARCH 模型实际上只适用于异方差函数短期自相关系数。但是有些残差序列的异方差函数是具有长期自关性,使用 ARCH 模型来做拟合,将会产生很高的移动平均阶数,增加参数估计的难度并影响拟合精度。为了解决这个问题,提出了广义自回归条件异方差模型,即 GARCH( $p, q$ )模型:

$$\sigma_i^2 = \alpha_0 + \alpha_1 \varepsilon_{i-1}^2 + \dots + \alpha_q \varepsilon_{i-q}^2 + \beta_1 \sigma_{i-1}^2 + \dots + \beta_p \sigma_{i-p}^2 \quad (2.6)$$

当  $p=0$  时, GARCH( $p, q$ )模型就是 ARCH 模型。

## 2.4. 指数平滑

指数平滑由布朗(Robert G. Brown)提出的一种在移动平均法的基础上发展而来的时间序列分析方法,配合以时间序列预测模型对未来的现象做出预测,通过加权平均法,对当前时段的观测值和预测值赋予不同的权重,从而得到下一时段的预测值,公式如下:

$$y'_{t+1} = a * y_t + (1-a) * y'_t \quad (2.7)$$

$a$  表示平滑系数;  $y'_{t+1}$ :  $t+1$  期的预测值,即本期( $t$  期)的平滑值;  $y_t$  是  $t$  期的实际值;  $y'_t$ :  $t$  期的预

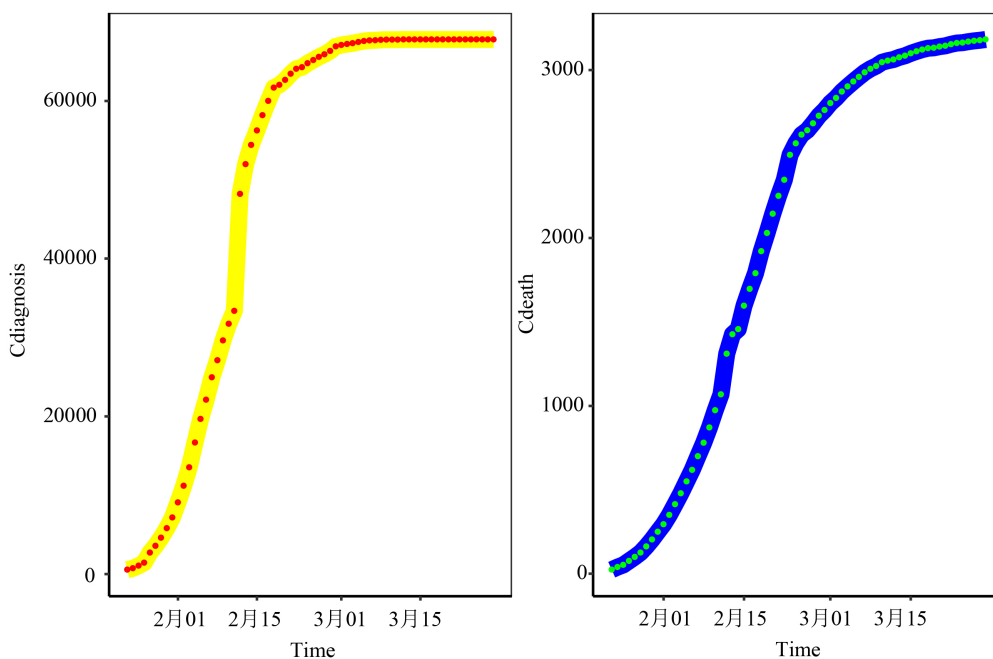
测值，即上期的平滑值。

指数平滑模型又根据时间序列是否存在长期趋势和季节效应分为简单指数平滑、Holt 两参数指数平滑、Holt-Winters 三参数数指数平滑，综合应用整体均值、整体趋势以及季节性进行预测，通过不同参数值的组合进行拟合。

### 3. 实证分析

#### 3.1. 数据描述

本节使用的数据集是关于湖北省 COVID-19 的累计死亡人数和累计确诊人数的数据集。我们在中华人民共和国国家卫生健康委员会上获取 2020 年 1 月 23 日至 2020 年 3 月 28 日的相关数据。其中 2020 年 1 月 23 日到 2020 年 3 月 21 日的数据用来训练模型，其余数据用来验证模型精度，如图 1 所示。由图 1 中我们可以看出在疫情初期累计确诊病例和累计死亡人数增长迅速，经过一段时间增长缓慢趋于平稳。



**Figure 1.** Time series of cumulative number of confirmed cases and cumulative number of deaths of COVID-19 in Hubei Province

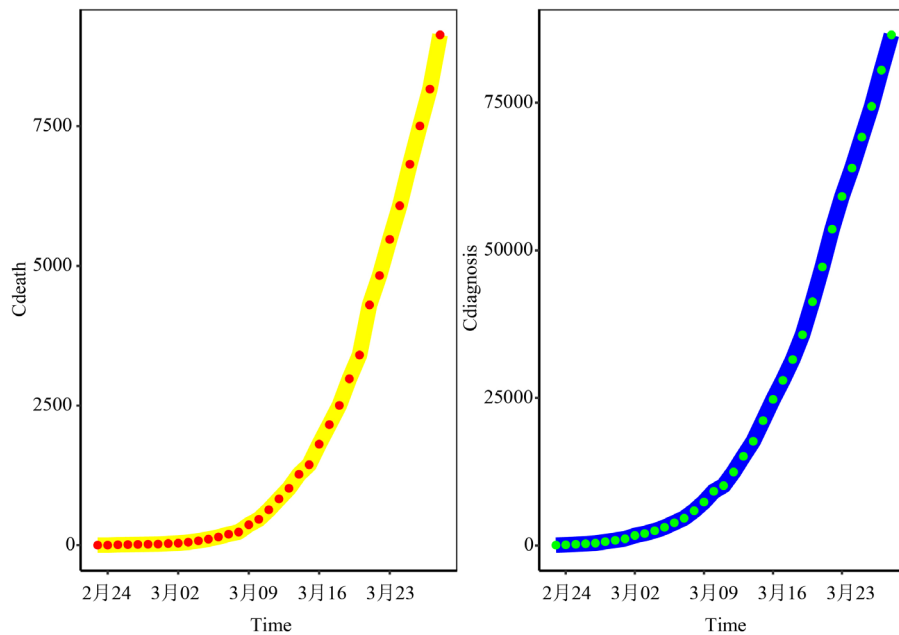
**图 1.** 湖北省累计确诊病例数和累计死亡病例数的时间序列

为了证明所提出的方法的可行性和适用性，我们选取意大利的累计确诊病例和累计死亡人数数据用来评估模型的性能。关于意大利 COVID-19 的数据集从 WHO 网址上获得。数据包含 2020 年 2 月 23 日到 2020 年 3 月 28 日的每日数据，其中 2020 年 2 月 23 日到 2020 年 3 月 21 日的日数据用来拟合模型，其余数据为验证数据，如图 2 所示。

#### 3.2. ES-GARCH 模型预测湖北省疫情趋势

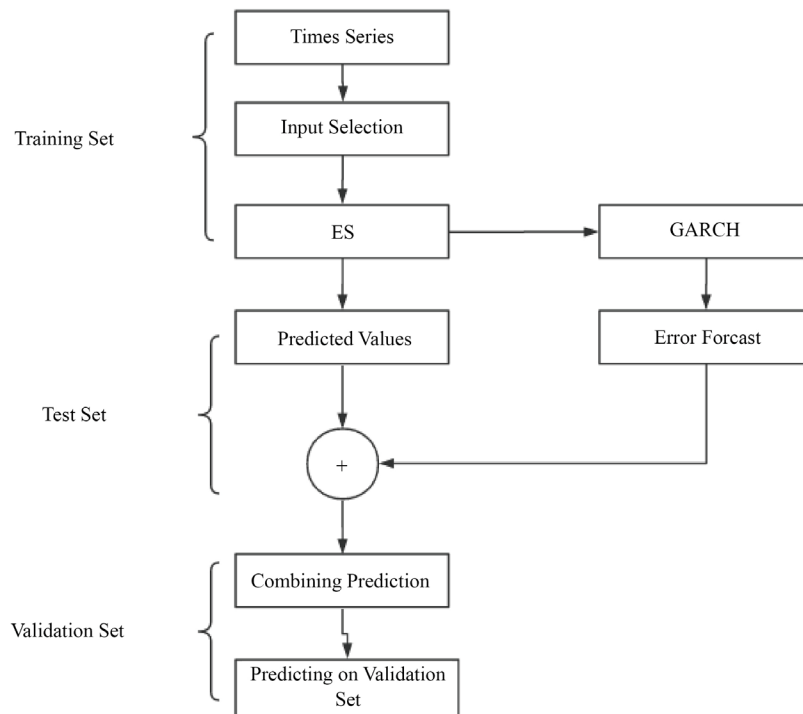
本部分以湖北省 2020 年 1 月 23 日至 2020 年 3 月 21 日的数据为训练集，建立数学模型模型。模型以 2020 年 3 月 22 日至 2020 年 3 月 28 日的日数据为验证集，验证了模型的准确性模型，目的是确定最优模型，预测湖北省累计死亡人数和累计确诊病例数省。本文的研究思路如图 3 所示。从表 1 和表 2 可

以看出, ES-GARCH 模型的模型性能与其他六个模型进行了比较。结果表明, ES-GARCH 模型的性能优于其他模型。



**Figure 2.** Time series of cumulative number of confirmed cases and cumulative number of deaths of COVID-19 in Italy

**图 2.** 意大利 COVID-19 累计确诊病例数和累计死亡人数的时间序列



**Figure 3.** The process of predicting confirmed cases and death toll using the ES-GARCH hybrid model

**图 3.** 使用 ES-GARCH 混合模型预测确诊病例和死亡人数的过程

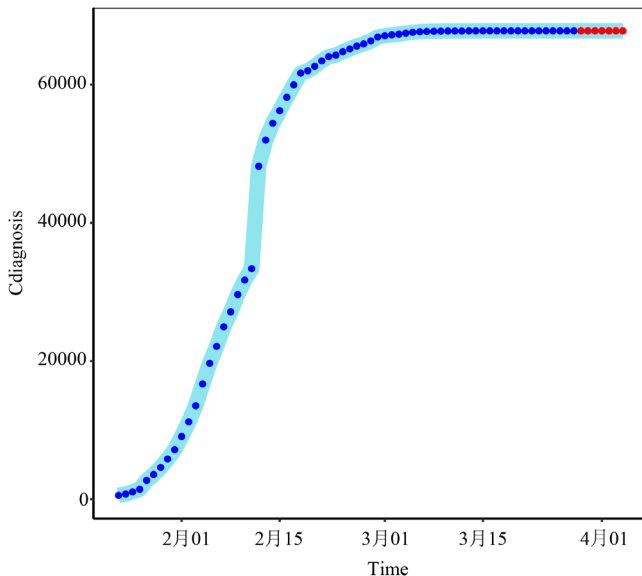
**Table 1.** Performance evaluation of the model using the cumulative number of confirmed cases data set in Hubei province  
**表 1.** 利用湖北省累计确诊病例数据集对模型进行性能评价

方法	MAE	MAPE (%)	RMSE
MA	42.5456	0.0628	44.2574
MA-SVR	5.9076	0.0087	5.9253
MA-GARCH	38.5561	0.0569	43.0524
MA-GARCH-SVR	9.8627	0.0145	10.0451
ES	0.8571	0.0013	0.9258
<b>ES-GARCH</b>	<b>0.4367</b>	<b>0.0006</b>	<b>0.5149</b>
ES-SVR	9.0021	0.0133	9.0696

**Table 2.** Performance evaluation of the model using the cumulative number of confirmed deaths data set in Hubei province  
**表 2.** 利用湖北省累计确认死亡数据对模型进行性能评价

方法	MAE	MAPE (%)	RMSE
MA	33.3909	1.0529	35.1824
MA-SVR	4.9571	0.1565	5.1441
MA-GARCH	31.3731	0.9890	34.0606
MA-GARCH-SVR	5.8524	0.1847	6.1423
ES	4.0909	0.1292	4.2329
<b>ES-GARCH</b>	<b>2.4079</b>	<b>0.0760</b>	<b>2.8816</b>
ES-SVR	2.4662	0.0779	3.0163

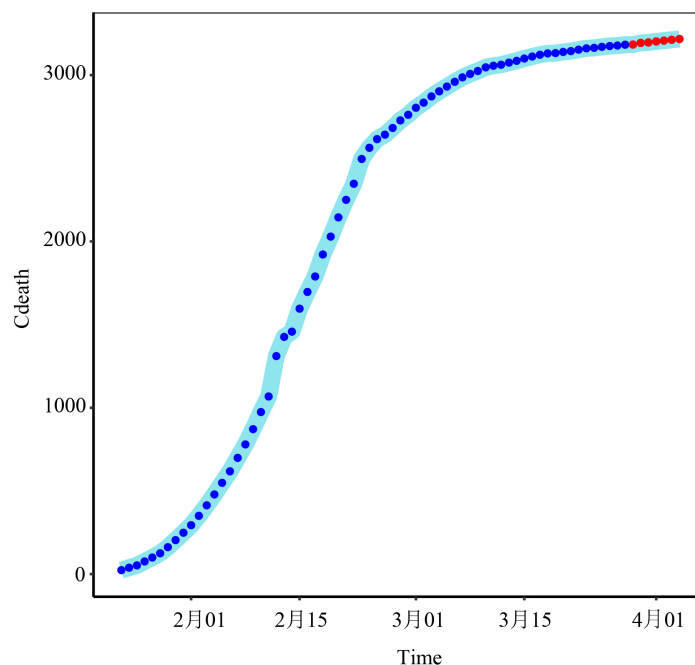
使用 ES-GARCH 模型可用于预测未来 7 天湖北省确诊病例和死亡病例。图 4、图 5 直观地显示了模型拟合数据的结果，以及未来几天疫情的变化趋势。我们可以直观地看到，未来几天湖北疫情趋于平稳，累计确诊病例数和累计死亡人数达到高峰。



注：蓝点表示 2020 年 1 月 23 日至 2020 年 3 月 28 日的真实值，红点表示 2020 年 3 月 29 日至 2020 年 4 月 4 日的预测值。

**Figure 4.** The cumulative number of confirmed cases in Hubei  
**图 4.** 湖北省累计确诊病例总数





注：蓝点表示 2020 年 1 月 23 日至 2020 年 3 月 28 日的真实值，红点表示 2020 年 3 月 29 日至 2020 年 4 月 4 日的预测值。

**Figure 5.** Total death toll in Hubei Province  
**图 5.** 湖北省累计死亡总人数

表 3 显示了使用 ES-GARCH 模型预测的 2020 年 3 月 29 日至 2020 年 4 月 4 日湖北省累计确诊病例数和累计死亡数。湖北省疫情已达高峰，此后确诊病例数和累计死亡数趋于平稳。由此可见，湖北省采取的防疫措施对控制疫情发展起到了积极作用，感染者得到了有效救治。

**Table 3.** The ES-GARCH model was used to predict confirmed cases and cumulative deaths of COVID-19  
**表 3.** 使用 ES-GARCH 模型预测新冠肺炎确诊病例和累计死亡病例

Date	确诊病例	累计死亡病例
29/3/2020	67801	3184
30/3/2020	67801	3194
31/3/2020	67801	3196
1/4/2020	67801	3202
2/4/2020	67801	3207
3/4/2020	67801	3212
4/4/2020	67801	3217

### 3.3. ES-GARCH 模型预测意大利疫情趋势

在本节中，我们将使用来自意大利的 COVID-19 数据来验证我们使用的 7 个模型的性能。表 4 和表 5 显示了意大利用于评估每个模型预测累积确诊病例和累积死亡人数的准确性的三个评估指标。我们可以直观地看到，在所有的模型中，只有 ES-GARCH 模型使三个评价指标最低，即 ES-GARCH 模型的性能最好。这一结论与上一节的结论是一致的。ES-GARCH 模型的预测精度最高。



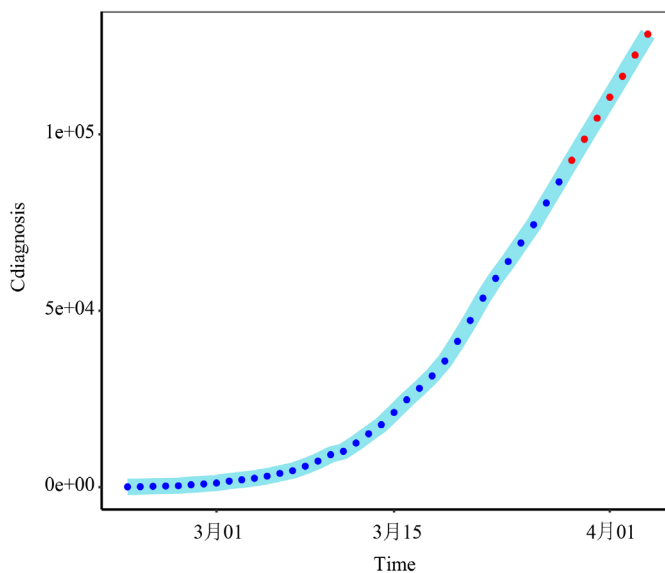
**Table 4.** Performance evaluation of the model using the cumulative number of confirmed cases data set in Italy.  
**表 4.** 利用意大利累计确诊病例数据集对模型进行绩效评价

方法	MAE	MAPE (%)	RMSE
MA	6685.370349	9.7044	6716.326859
MA-SVR	463.2821929	0.6802	562.2130918
MA-GARCH	5037.557143	7.6718	5200.69408
MA-GARCH-SVR	8230.228571	11.7432	8400.83351
ES	1358.183167	1.8394	1562.838369
<b>ES-GARCH</b>	<b>67.70714286</b>	<b>0.1014</b>	<b>68.35630716</b>
ES-SVR	422.8252604	3.9861	510.5498386

**Table 5.** Performance evaluation of the model using the cumulative number of confirmed deaths data set in Italy  
**表 5.** 利用意大利的累计确认死亡数据对模型进行性能评价

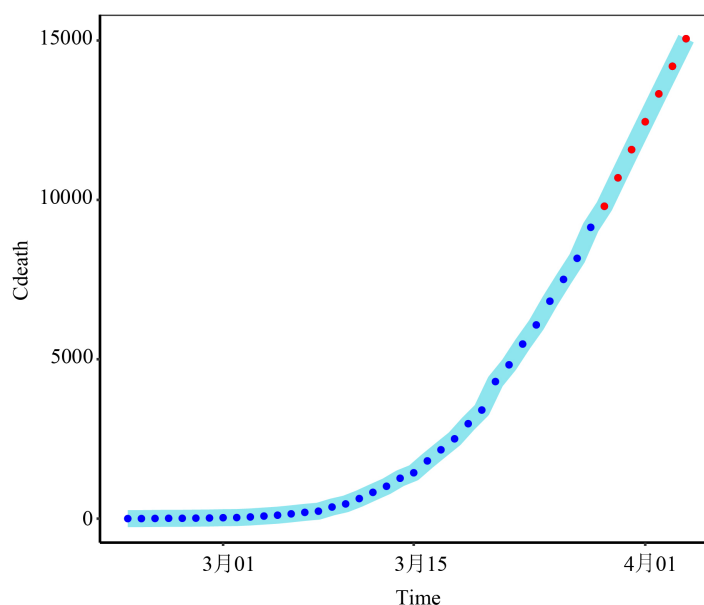
方法	MAE	MAPE (%)	RMSE
MA	1931.1719	27.1611	2061.9339
MA-SVR	434.8900	6.0413	528.0000
MA-GARCH	555.7286	7.5378	639.1753
MA-GARCH-SVR	1373.0971	22.6836	1487.5791
ES	461.1795	6.5007	495.1102
<b>ES-GARCH</b>	<b>460.2901</b>	<b>6.4138</b>	<b>492.2376</b>
ES-SVR	2784.2381	5.8556	3543.6219

图 6 和图 7 分别为 2020 年 3 月 29 日至 4 月 4 日意大利累计确诊病例和累计死亡病例的趋势，采用 ES-GARCH 模型。由图 6 可知，意大利确诊病例数在未来 7 天内将快速增加。死亡人数证实，这种疾病的迅速蔓延给意大利的卫生系统带来了新的挑战。



注：蓝点表示 2020 年 2 月 23 日至 2020 年 3 月 28 日的真实值，红点表示 2020 年 3 月 29 日至 2020 年 4 月 4 日的预测值。

**Figure 6.** The total number of confirmed cases in Italy  
**图 6.** 意大利累计确诊病例总数



注：蓝点表示 2020 年 2 月 23 日至 2020 年 3 月 28 日的真实值，红点表示 2020 年 3 月 29 日至 2020 年 4 月 4 日的预测值。

**Figure 7.** The cumulative number of confirmed deaths in Italy  
**图 7.** 意大利确认的累计死亡总数

#### 4. 结论

新型冠状病毒的爆发严重影响了人们的正常生活，对人民的身体健康造成了极大的伤害。因此，针对疫情发展趋势，采取有效的防护措施是至关重要的。利用本文所提出 ES-GARCH 模型可以作为一种预测 COVID-19 的预测方法，为制定决策提供可靠依据。ES-GARCH 模型对未来七天内确诊病例数和累计死亡数具有较高的预测能力。同时 RMSE、MAE、MAPE 等评价指标也为我们确定最优模型提供有力的依据。利用意大利确诊病例和累计死亡数据对该模型进行了验证，评价结果表明该方法具有良好的效果。

从本文研究中可以得到结论，使用 ES-GARCH 模型预测结果较为准确。结合预测结果我们发现，湖北省的确诊病例和死亡人数达到了一个相对平稳的趋势。综合此模型的各项评价指标，可以得出结论，该模型可以很好拟合短期数据，并可以实现相对精准的短期预测。ES-GARCH 模型可以为决策者提供了一个有力的预测工具，根据历史数据以及未来疫情的发展趋势，可以做到及时防护和采取相应措施。

#### 参考文献

- [1] Wu, J.T., Leung, K. and Leung, G.M. (2020) Nowcasting and Forecasting the Potential Domestic and International Spread of the 2019-nCoV Outbreak Originating in Wuhan, China: A Modelling Study. *The Lancet*, **395**, 689-697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)
- [2] Wang, L.F., Shi, Z., Zhang, S., Field, H., Daszak, P. and Eaton, B.T. (2006) Review of Bats and SARS. *Emerging Infectious Diseases*, **12**, 1834-1840. <https://doi.org/10.3201/eid1212.060401>
- [3] Riou, J. and Althaus, C.L. (2020). Pattern of Early Human-to-Human Transmission of Wuhan 2019-nCoV. [Preprint]
- [4] World Health Organization (2003) Consensus Document on the Epidemiology of Severe Acute Respiratory Syndrome (SARS). World Health Organization, Geneva. <https://who.int/csr/sars/en/WHOconsensus.pdf>
- [5] Steinbuch, Y. (2020) China Opens Second New Hospital for Coronavirus Patients.
- [6] Cheng, Z.J. and Shan, J. (2020) 2019 Novel Coronavirus: Where We Are and What We Know. *Infection*, **48**, 155-163. <https://doi.org/10.20944/preprints202001.0381.v1>
- [7] Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S. and Ciccozzi, M. (2020) Application of the ARIMA Model on the COVID-2019 Epidemic Dataset. *Data in Brief*, **29**, Article ID: 105340.

- <https://doi.org/10.1016/j.dib.2020.105340>
- [8] Hu, Z., Ge, Q., Li, S., Jin, L. and Xiong, M. (2020) Artificial Intelligence Forecasting of Covid-19 in China.
- [9] Weber, A., Iannelli, F. and Gonçalves, S. (2020) Trend Analysis of the COVID-19 Pandemic in China and the Rest of the World. *medRxiv*. <https://doi.org/10.1101/2020.03.19.20037192>
- [10] Jung, S.M., Akhmetzhanov, A.R., Hayashi, K., Linton, N.M., Yang, Y., Yuan, B., Kobayashi, T., Kinoshita, R. and Nishiura, H. (2020) Real Time Estimation of the Risk of Death from Novel Coronavirus (2019-nCoV) Infection: Inference Using Exported Cases. *Journal of Clinical Medicine*, **9**, Article No. 523. <https://doi.org/10.3390/jcm9020523>
- [11] Zhao, S., Musa, S.S., Lin, Q., Ran, J., Yang, G., Wang, W., *et al.* (2020) Estimating the Unreported Number of Novel Coronavirus (2019-nCoV) Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak. *Journal of Clinical Medicine*, **9**, Article No. 388. <https://doi.org/10.3390/jcm9020388>
- [12] Guan, P., Wu, W. and Huang, D. (2018) Trends of Reported Human Brucellosis Cases in Mainland China from 2007 to 2017: An Exponential Smoothing Time Series Analysis. *Environmental Health and Preventive Medicine*, **23**, Article No. 23. <https://doi.org/10.1186/s12199-018-0712-5>
- [13] Wang, S., Petzold, M., Cao, J., Zhang, Y. and Wang, W. (2015) Direct Medical Costs of Hospitalizations for Cardiovascular Diseases in Shanghai, China. *Medicine*, **94**, e837. <https://doi.org/10.1097/MD.0000000000000837>
- [14] Zhou, L., Zhao, P., Wu, D., Cheng, C. and Huang, H. (2018) Time Series Model for Forecasting the Number of New Admission Inpatients. *BMC Medical Informatics and Decision Making*, **18**, Article No. 39. <https://doi.org/10.1186/s12911-018-0616-8>
- [15] Leneenadogo, W. and Godwin Lebari, T. (2019) Modelling the Nigeria Crude Oil Prices Using ARIMA, Pre-Intervention and Post-Intervention Model. *Asian Journal of Probability and Statistics*, **3**, 1-12. <https://doi.org/10.9734/ajpas/2019/v3i130083>
- [16] Mao, Q., Zhang, K., Yan, W. and Cheng, C. (2018) Forecasting the Incidence of Tuberculosis in China Using the Seasonal Auto-Regressive Integrated Moving Average (SARIMA) Model. *Journal of Infection and Public Health*, **11**, 707-712. <https://doi.org/10.1016/j.jiph.2018.04.009>
- [17] Xu, Q., Li, R., Liu, Y., Luo, C., Xu, A., Xue, F., *et al.* (2017) Forecasting the Incidence of Mumps in Zibo City Based on a SARIMA Model. *International Journal of Environmental Research and Public Health*, **14**, Article No. 925. <https://doi.org/10.3390/ijerph14080925>
- [18] Zhang, X., Zhang, T., Pei, J., Liu, Y., Li, X., Medrano Gracia, P., *et al.* (2016) Time Series Modelling of Syphilis Incidence in China from 2005 to 2012. *PLoS ONE*, **11**, e0149401. <https://doi.org/10.1371/journal.pone.0149401>
- [19] Earnest, A., Chen, M.I., Ng, D. and Sin, L.Y. (2005) Using Autoregressive Integrated Moving Average (ARIMA) Models to Predict and Monitor the Number of Beds Occupied during a SARS Outbreak in a Tertiary Hospital in Singapore. *BMC Health Services Research*, **5**, Article No. 36. <https://doi.org/10.1186/1472-6963-5-36>
- [20] Yang, Y., Zhang, X., Zhang, T., Young, A.A. and Li, X. (2014) Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data. *PLoS ONE*, **9**, e91629. <https://doi.org/10.1371/journal.pone.0088075>