International Journal of Systems Science and Applied Mathematics
2020; 5(2): 20-26
http://www.sciencepublishinggroup.com/j/ijssam
doi: 10.11648/j.ijssam.20200502.12
ISSN: 2575-5838 (Print); ISSN: 2575-5803 (Online)

Covid-19 Projections: Single Forecast Model Against Multi-Model Ensemble

Otoo Joseph[1], Bosson-Amedenu Senyefia[2, *], Nyarko Christiana Cynthia[2], Osei-Asibey Eunice[2], Boateng Ernest Yeboah[3]

[1]Department of Statistics and Actuarial Science, University of Ghana, Legon, Accra, Ghana

[2]Department of Mathematical Sciences, University of Mines and Technology, Tarkwa, Ghana

[3]Department of Basic Sciences, School of Basic and Biomedical Sciences, University of Health and Allied Sciences, Ho, Ghana

Email address:

sbosson-amedenu@st.umat.edu.gh (Bosson-Amedenu S.)

[*]Corresponding author

To cite this article:

Otoo Joseph, Bosson-Amedenu Senyefia, Nyarko Christiana Cynthia, Osei-Asibey Eunice, Boateng Ernest Yeboah. Covid-19 Projections: Single Forecast Model Against Multi-Model Ensemble. *International Journal of Systems Science and Applied Mathematics*. Vol. 5, No. 2, 2020, pp. 20-26. doi: 10.11648/j.ijssam.20200502.12

Received: July 4, 2020; **Accepted**: July 20, 2020; **Published**: July 28, 2020

Abstract: The novel coronavirus has unsettled many nations and has created severe uncertainty in its spread. In this paper, we present the performance of ensemble models and single forecast models in the projection of COVID-19 confirmed cases in nine countries. Data consisting of two (2) health indicators (new COVID-19 and cumulative COVID-19 confirmed cases) were collated on May 10, 2020 from the Humanitarian Data Exchange (HDX). Forecasting models with the minimum Mean Square Error (MSE) and Root Mean Square Error (RMSE) were selected. Our findings showed that ETS (A, N, N) was the best model fit for China, Spain, South Korea and Ghana in terms of single COVID-19 confirmed cases. On the other hand, INGARCH (1, 1) was the best fit model for the remaining countries. Regarding cumulative COVID-19 confirmed cases, INGARCH (1, 1) was fit for each of the nine countries. Again, we found that single forecasting models outperform hybrid models when the number of data points does not meet a certain threshold, and when the data has no seasonality; suggesting further that hybrid forecast models perform efficiently in complex time series dataset. Results from the 10 days forecast indicate that for most countries, with the exception of Ghana and India, new covid-19 confirmed cases will drop. The study suggest for future works to expand the training dataset by augmenting additional data onto the available data and then apply hybrid forecasting models to the dataset.

Keywords: COVID-19, Coronavirus, Ensemble, Forecasting, Multi-Model, Time Series

1. Introduction

The World Health Organization (WHO) declared COVID-19 outbreak a pandemic on March 12, 2020; after it had spread widely across China (since December 2019), extending into fifty-one (51) countries by February 28, 2020 [1]. The declaration of International Public Health Emergency concern was a form of alert for countries to be prepared for containment through active surveillance, early detection, isolation and case management, contact tracing; to enhance prevention and forestall further spread of the novel coronavirus [2].

The influx of vertical (imported cases) and horizontal (local transmissions) spread of COVID-19 pandemic has necessitated several efforts by researchers to forecast the morbidity, mortality cases and recoveries using varied models. As a result, many models have been proposed which have suggested diverse contexts of forecasts towards unique needs. Most of these proposed forecast models were developed at the early stages of the pandemic, where the number of cases did not encourage holding a substantial amount as training data to allow for efficient estimates. Researchers agree for the need to explore forecasting methods and tools to promote accurate predictions [3].

Li et al., [1] in their quest to predict the spread of COVID-19 in China developed ARIMAX (0, 1, 0) model with R-square value of 0.977 and a corresponding Ljung-Box Q (18) test statistic value of 0.987. The researchers explained from their results that China's emergency intervention measures at the onset of the epidemic had a critical restraining effect on the original spread of the epidemic.

Dehesh et al., [4] also sought to forecast confirmed cases of COVID-19 in different Countries using ARIMA Models. Their finding suggested the models ARIMA (2,1,0), ARIMA (2,2,2) and ARIMA (1,0,0) as respective robust models for Mainland China, Italy and South Korea. Other suggested models were ARIMA (2,3,0) and ARIMA (3,1,0) for Iran and China respectively.

Zhang, [5] found from his use of hybrid model in forecasting that the combined model approach has great potential to improving forecasting accuracy than when either of the models was used separately. Wang, [6] supported this view when he had similar findings that suggested that using hybrid models provided improved performances in forecasting.

In this paper, we aim at comparing projections of COVID-19 confirmed cases (both single and cumulative) using single forecast model and multi-model ensemble for nine countries. These countries included China, India, Iran, Italy, Spain, Thailand, Turkey, South Korea and Ghana.

2. Materials and Methods

Data

Data were collated on May 10, 2020 from the Humanitarian Data Exchange (HDX)[1]. The data spans from January 11, 2020 to May 10, 2020. The data consist of two (2) health indicators such as new covid-19 confirmed cases and cumulative covid-19 confirmed cases for nine (9) countries. The countries were China, India, Iran, Italy, Spain, Thailand, Turkey, South Korea and Ghana.

The dataset was divided into training and testing dataset. The composition of the training and testing dataset were 80% and 20% respectively.

Figure 1 shows the time series plot of daily new infections/cases against time for India, Iran, Italy, Spain and Ghana. Clearly from the figure there were no consistent trends (upward or downward) over the entire time span for all the countries. The series appeared to wander up quickly to its peak and wandered down slowly for countries like Spain, Italy and Iran. In Ghana and India, the series were wandering up but had not reached their peak yet. There were no seasonality and no obvious outliers identified for all the countries.

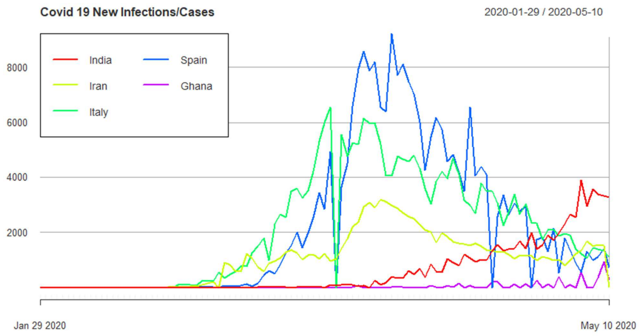

Figure 1. *Time Series Plot of Covid-19 New Infections/Cases for India, Iran, Italy, Spain and Ghana.*

Figure 2 shows the time series plot of daily new infections/cases against time for China, South Korea, Thailand and Turkey. These series showed no consistent trends (upward or downward) over the entire time span for all the countries. The series appeared to wander up quickly to its peak and wander down slowly for Turkey than the other countries. There was no seasonality identified for all countries but an obvious outlier for China.

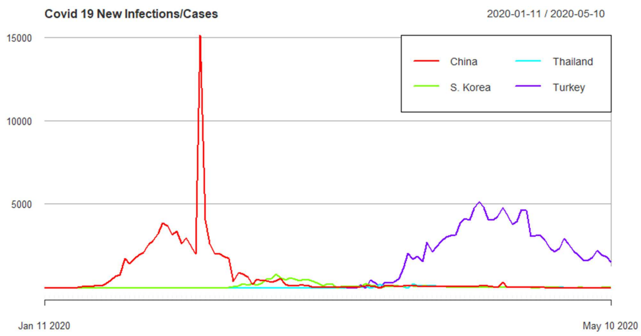

Figure 2. *Time Series Plot of Covid-19 New Infections/Cases for China, Thailand, South Korea and Turkey.*

Model Estimation

We compared several single forecasting models with ensemble/hybrid forecasting models. Due to limitation of space, forecasting models with similar or close forecasting accuracy metrics were covered in this study. In this regard, we compared the ARIMA forecasting model, the Theta forecasting model, Simple Exponential Smoothing and the Integer Valued GARCH model.

We developed several potential time series models based on the techniques stated in the foregoing paragraph using 80% of the dataset of the selected countries. The developed time series model were then used to forecast based on the length (the number of observation) of the remaining 20% of the dataset. We then estimated the Mean Square Error (MSE) and the Root Mean Square Error (RMSE) between the forecasted values and the remaining 20% of the dataset. Forecasting models with the minimum MSE and RMSE were selected. We also forecasted ten (10) days of new covid-19 cases for the nine (9) countries using the selected models.

Mathematical Formulation

1 https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases

ETS

Considering an observed time series: $y_1, y_2, \ldots, y_n$. The Simple Exponential Smoothing equation formally takes the form:

$$\hat{y}_{i+1} = \alpha y_i + (1-\alpha)\hat{y}_i \quad (1)$$

Where y_i is the actual, known series value for time period i, $\hat{y}_i$ is the forecast value of the variable Y for time period i, $\hat{y}_{i+1}$ is the forecast value for time period $i+1$ and α is the smoothing constant [7].

The forecast $\hat{y}_{i+1}$is based on weighting the most recent observation y_i with a weight α and weighting the most recent forecast $\hat{y}_i$ with a weight of $1-\alpha$. For details of how this forecasting method works see [7].

INGARCH

Ferland et al., [8] and Fokianos et al., [9] proposed the INGARCH model which is stated as follows:

$$\begin{cases} X_t|\mathcal{F}_{t-1} \sim \mathbb{P}(\lambda_t), \\ \lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \lambda_{t-1} + \sum_{j=1}^{q} \beta_j X_{t-j}, \end{cases} \quad t \in \mathbb{Z}, \quad (2)$$

Where $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, i = 1, \ldots, p, j = 1, \ldots., q, p \geq 0, q \geq 1$, and $\mathcal{F}_{t-1}$ is the σ −field [10].

3. Results

The new covid-19 confirmed cases for the period of January 11, 2020 to May 10; 2020 exhibited no consistent trends (upward or downward) over the entire time span for all the countries. The series appeared to wander up quickly to its peak and wandered down slowly for countries like Spain, Italy, Iran and Turkey. Ghana and India had their series (new covid-19 confirmed cases) wandered up and yet to hit its peak. There was no seasonality identified for all countries but an obvious outlier for China which concurs with findings from the study of [11]. These results are well captured by Figures 1 and 2.

Among the fitted time series forecasting models for new covid-19 confirmed cases and cumulative covid-19 confirmed cases (Tables 1 and 2), single forecasting models such as ETS and INGARCH models outperformed ARIMA and the Hybrid forecast model of ARIMA and Theta model. Due to insufficient amount of available data for the novel covid-19 and the non-seasonality of the dataset, several models single forecasting models and multi-model ensemble models could not be fitted. For instance, Seasonal and Trend decomposition using Loess models require that the input series be seasonal; furthermore, the data must include at least two seasons of data for the decomposition to succeed. Similarly, Neural Network Time Series Forecasts models also require that the data must include at least two seasons of data.

Several studies [12-14] have shown the efficiency of the hybrid forecasting models in improving forecast accuracy. However, due to insufficient amount of available data for the novel covid-19 and the non-seasonality of the dataset, the single forecasting models outperformed the hybrid or multi-model ensemble models as evidenced by Tables 1 and 2. This results also suggest that hybrid forecast models performs efficiently of complex time series dataset [13, 14].

There are several ways to build a most accurate forecasting model with limited or insufficient dataset. Three of these approaches were expounded by [12]. We suggest for future works or studies to expand the training dataset by augmenting additional data onto the available data and the application of the hybrid forecasting models to the dataset.

The selected models together with its estimated parameters for new covid-19 confirmed cases and cumulative covid-19 confirmed cases for the nine (9) countries are well captured by Table 3.

Results from the 10 days forecast indicated a downward trend of new covid-19 confirmed cases for most of the countries considered in this study; with the exception of Ghana and India. For Ghana and India, new covid-19 confirmed cases will rise or increase as captured by the 10 days forecast.

Table 1. *Comparison of Single forecasting Models and Multi-model Ensemble for New Covid-19 Cases.*

Model	Country	Metrics	
		RMSE	MSE
HYBRID (ARIMA + TBATS)	China	56054.45	3142101127
ETS	China	137.16	18812.91
INGARCH	China	446.95	199764.6
HYBRID (ARIMA + TBATS)	India	13010.38	169269925
ETS	India	872.41	761102.5
ARIMA	India	1211.26	1467145
INGARCH	India	138.41	19157.38
HYBRID (ARIMA + TBATS)	Iran	239607.1	57411584184
ETS	Iran	464.66	215908.1
ARIMA	Iran	454.76	206806.7
INGARCH	Iran	358.98	128865.2
HYBRID (ARIMA + TBATS)	Italy	13574.81	184275341
ETS	Italy	1199.71	1439310
ARIMA	Italy	1303.10	1698068
INGARCH	Italy	1116.44	1246439
HYBRID (ARIMA + TBATS)	Spain	35012.75	1225892802
ETS	Spain	486.73	236907.4

Model	Country	Metrics	
		RMSE	MSE
ARIMA	Spain	543.02	294872.1
INGARCH	Spain	641.85	411976.4
HYBRID (ARIMA + TBATS)	South Korea	1525422	2.326911e+12
ETS	South Korea	10.90	118.75
ARIMA	South Korea	54.84	3007.14
INGARCH	South Korea	10.99	120.69
HYBRID (ARIMA + TBATS)	Thailand	1927.50	3715249
ETS	Thailand	11.14	124.18
ARIMA	Thailand	14.68	215.42
INGARCH	Thailand	2.60	6.77
HYBRID (ARIMA + TBATS)	Turkey	204364.4	41764816826
ETS	Turkey	860.69	740785.1
ARIMA	Turkey	863.40	745453.3
INGARCH	Turkey	98.10	9623.25
HYBRID (ARIMA + TBATS)	Ghana	304.97	93009
ETS	Ghana	290.82	84576.89
ARIMA	Ghana	323.05	104361.4
INGARCH	Ghana	508.29	258357.1

Table 2. *Comparison of Single forecasting Models and Multi-model Ensemble for Cumulative Covid-19 Cases.*

Model	Country	Metrics	
		RMSE	MSE
HYBRID (ARIMA + TBATS)	China	18079.2	326857352
ARIMA	China	19563.39	382726222
INGARCH	China	12439.01	154729002
HYBRID (ARIMA + TBATS)	India	13475.36	181585307
ARIMA	India	67955.76	4617985365
INGARCH	India	3173.72	10072481
HYBRID (ARIMA + TBATS)	Iran	24172.99	584333417
ARIMA	Iran	27053.14	731872450
INGARCH	Iran	19760.9	390493260
HYBRID (ARIMA + TBATS)	Italy	57585.26	3316062000
ARIMA	Italy	1053172	1.10917e+12
INGARCH	Italy	21457.21	460411796
HYBRID (ARIMA + TBATS)	Spain	59912.18	3589469125
ARIMA	Spain	613925.5	376904496141
INGARCH	Spain	24912.18	620616712.35
HYBRID (ARIMA + TBATS)	South Korea	2157.568	4655098
ARIMA	South Korea	2327.052	5415173
INGARCH	South Korea	1334.78	1781646
HYBRID (ARIMA + TBATS)	Thailand	705.81	498167.2
ARIMA	Thailand	7813.504	61050840
INGARCH	Thailand	86.74	7523.112
HYBRID (ARIMA + TBATS)	Turkey	41523.95	1724238694
ARIMA	Turkey	41700.13	1738901237
INGARCH	Turkey	27054.8	731962110
HYBRID (ARIMA + TBATS)	Ghana	1701.52	2895156
ARIMA	Ghana	1673.43	2800351
INGARCH	Ghana	735.47	540916.2

Table 3. *Selected Forecasting Models for New and Cumulative Covid-19 Cases.*

Country	Selected Model	Estimated Parameters
New Covid-19 Cases		
China	ETS(A,N,N)	$\alpha = 0.30; l = 23.22$
India	INGARCH	$\alpha = 51.7; \beta_{t-1} = 0.89$
Iran	INGARCH(1,1)	$\alpha = 1050; \beta_{t-1} = 0.38$
Italy	INGARCH(1,1)	$\alpha = 2160; \beta_{t-1} = 0.40$
Spain	ETS(A,N,N)	$\alpha = 0.58; l = 0.30$
South Korea	ETS(A,N,N)	$\alpha = 0.91; l = 0.59$
Thailand	INGARCH(1,1)	$\alpha = 0.04; \beta_{t-1} = 0.35;\ \alpha_{t-1} = 0.64$
Turkey	INGARCH(1,1)	$\alpha = 1790; \beta_{t-1} = 0.44$
Ghana	ETS(A,A,N)	$\alpha = 0.01; \beta = 0.01; l = 2.56; b = -0.25$
Cumulative Covid-19 Cases		
China	INGARCH(1,1)	$\alpha = 56800; \beta_{t-1} = 0.18$
India	INGARCH(1,1)	$\alpha = 678; \beta_{t-1} = 0.87$

Country	Selected Model	Estimated Parameters
New Covid-19 Cases		
Iran	INGARCH(1,1)	$a = 33800; \beta_{t-1} = 0.41$
Italy	INGARCH(1,1)	$a = 48900; \beta_{t-1} = 0.62$
Spain	INGARCH(1,1)	$a = 47100; \beta_{t-1} = 0.66$
South Korea	INGARCH(1,1)	$a = 4810; \beta_{t-1} = 0.42$
Thailand	INGARCH(1,1)	$a = 18.68; \beta_{t-1} = 0.97;\ \alpha_{t-1} = 0.004$
Turkey	INGARCH(1,1)	$a = 38100;\ \beta_{t-1} = 0.46$
Ghana	INGARCH(1,1)	$a = 108;\ \beta_{t-1} = 0.87$

NB: The parameter a in INGARCH (1, 1) is the intercept. The parameter l and b are the initial state.

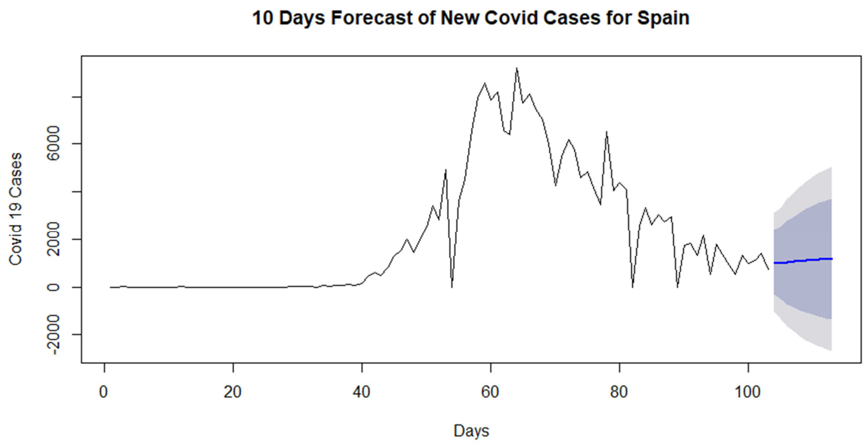

Figure 3. *Ten (10) days forecast of New Covid-19 cases for Spain.*

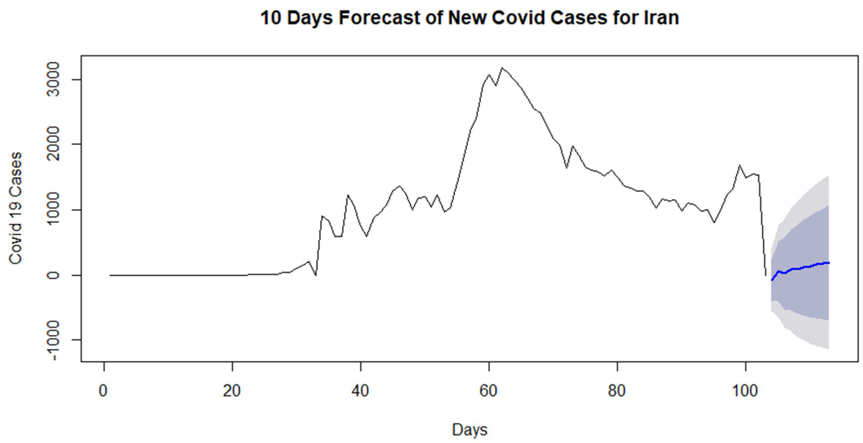

Figure 4. *Ten (10) days forecast of New Covid-19 cases for Iran.*

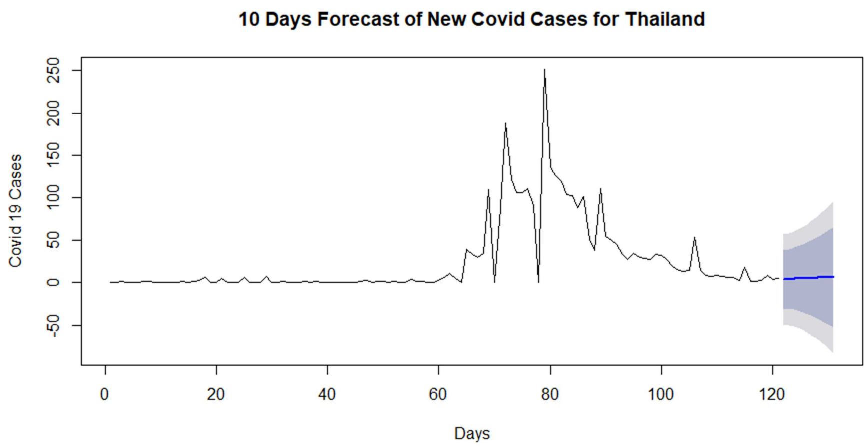

Figure 5. *Ten (10) days forecast of New Covid-19 cases for Thailand.*

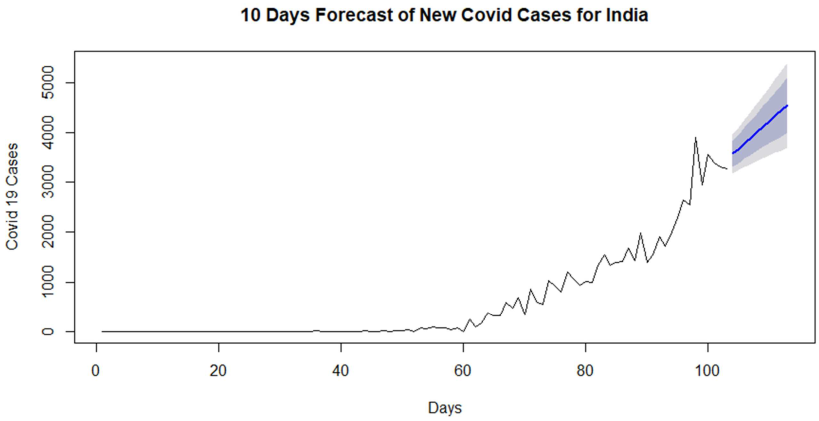

Figure 6. *Ten (10) days forecast of New Covid-19 cases for India.*

4. Discussion

Our findings showed that with the improvement in the size and nature of data counts, INGARCH (1,1) was the best model fit for countries such as China, Iran, South Korea, and Italy among others which hitherto had ARIMA models as their best fit models for the COVD-19 confirmed cases as demonstrated by in the work of Dehesh *et al*., [4]. Our finding provide scientific evidence for the need for researchers to improve on forecasting models as the size of COVID-19 cases increase to provide efficient models. The INGARCH (1, 1) was Ghana's best fit model since the pandemic, as no prior model has been formulated for Ghana. Our results suggest that Ghana which happens to be among the top three countries in Africa with high COVID-19 cases is yet to reach its peak. Results from the 10 days forecast indicate that most of the countries with the exception of Ghana and India new covid-19 confirmed cases going down. For Ghana and India, new covid-19 confirmed cases will rise or increase as captured by the 10 days forecast. It was interesting to note that countries such as India, Iran, Italy and Turkey had INGARCH (1, 1) as best model fit for both single and cumulative COVID-19 cases. The remaining countries either had ETS (A, N, N) or INGARCH (1, 1) for new COVID-19 confirmed and cumulative cases.

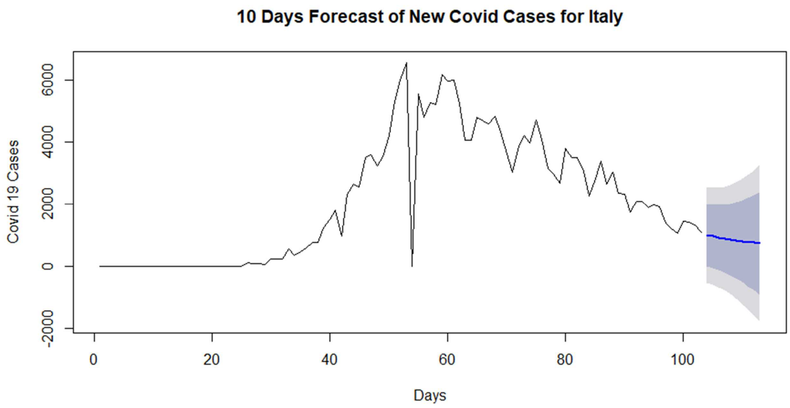

Figure 7. *Ten (10) days forecast of New Covid-19 cases for Italy.*

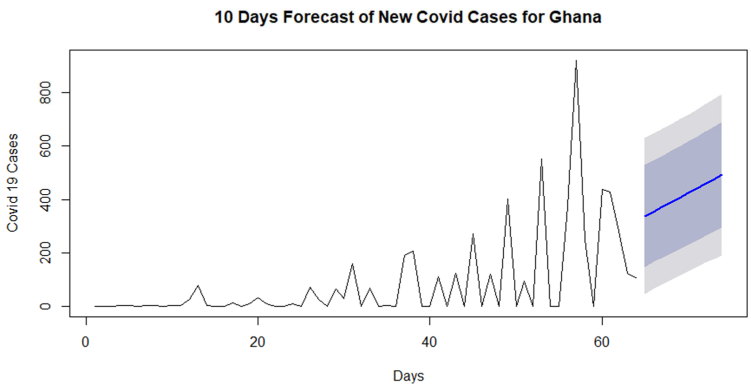

Figure 8. *Ten (10) days forecast of New Covid-19 cases for Ghana.*

5. Conclusion

In this study, we compared several single forecasting models with ensemble/hybrid forecasting models. Specifically, the ARIMA forecasting model, the Theta forecasting model, Simple Exponential Smoothing and the Integer Valued GARCH models were compared. We fitted time series forecasting models using 80% of the dataset for the nine (9) selected countries. The fitted time series model were used to forecast based on the length (the number of observation) of the remaining 20% of the dataset. Mean Square Error (MSE) and the Root Mean Square Error (RMSE) between the forecasted values and the remaining 20% of the dataset were estimated. Forecasting models with the minimum MSE and RMSE were selected. Ten (10) days of new covid-19 cases for the nine (9) countries were forecasted using the selected models.

6. Recommendation

We suggest for future works or studies to expand the training dataset by augmenting additional data onto the available data and the application of the hybrid forecasting models to the dataset. The study also suggested that hybrid forecast models should be used for complex time series dataset.

References

[1] Lai S., Ruktanonchai N. W, Zhou L., Prosper O., Luo W., Floyd J. R, Wesolowski A., Santillana M., Zhang C., Du X., Yu H., and Tatem A. J. Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. medRxiv preprint; 2020.

[2] Anon. National Contingency Plan for COVID-19, The Philippines; 2020.

[3] Luo J. Predictive Monitoring of COVID-19. Data-Driven Innovation Lab Singapore University of Technology and Design; 2020.

[4] Dehesh T. Mardani-Fard H. A, Dehesh P. Forecasting of COVID-19 Confirmed Cases in Different Countries with ARIMA Models. medRxiv preprint; 2020.

[5] Zhang, G. P. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175; 2020.

[6] Wang YH. Nonlinear neural network forecasting model for stock index option price: hybrid GJR-GARCH approach. Expert Systems with Applications 2009; 36: 564–70.

[7] Ostertagova, E., & Ostertag, O. Forecasting Using Simple Exponential Smoothing Method. Acta Electrotechnica et Informatica, 12 (3), 62-66; (2012).

[8] Ferland, R., Latour, A., & Oraichi, D. Integer-valued GARCH process. Journal of Time Series Analysis, 27, 923–942; (2006).

[9] Fokianos, K., Rahbek, A., & Tjøstheim, D. Poisson Autoregression. Journal of the American Statistical Association, 104, 1430–1439; (2009).

[10] Cui, Y., Li, Q., & Zhu, F. Flexible bivariate Poisson integer-valued GARCH model. Annals of the Institute of Statistical Mathematics. doi: 10.1007/s10463-019-00732-4; (2019).

[11] Fong, S. J., Li, G., Dey, N., & Crespo, R. G. (2020). Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak. International Journal of Interactive Multimedia and Artificial Intelligence, 6 (1), 132-140. doi: 10.9781/ijimai.2020.02.002.

[12] Araujo, M. B., & New, M. (2006). Ensemble forecasting of species distributions. *TRENDS in Ecology and Evolution, 22* (1), 42-47. doi: 10.1016/j.tree.2006.09.010.

[13] Hong, W.-C. (2009). Hybrid evolutionary algorithms in a SVR-based electric load forecasting model. *Electrical Power and Energy Systems, 31*, 409-417. doi: 10.1016/j.ijepes.2009.03.020.

[14] Xiao, L., Shao, W., Liang, T., & Wang, C. (2016). A combined model based on multiple seasonal patterns and modified firefly algorithm for electrical load forecasting. *Applied Energy, 167*, 135-153. doi: 10.1016/j.apenergy.2016.01.050.