ORIGINAL ARTICLE

Forecasting COVID-19 infections in the Arabian Gulf region

Sami Khedhiri[1] ⬤

Abstract

In this paper, an empirical analysis of linear state space models and long short-term memory neural networks is performed to compare the statistical performance of these models in predicting the spread of COVID-19 infections. Data on the pandemic daily infections from the Arabian Gulf countries from 2020/03/24 to 2021/05/20 are fitted to each model and a statistical analysis is conducted to assess their short-term prediction accuracy. The results show that state space model predictions are more accurate with notably smaller root mean square errors than the deep learning forecasting method. The results also indicate that the poorer forecast performance of long short-term memory neural networks occurs in particular when health surveillance data are characterized by high fluctuations of the daily infection records and frequent occurrences of abrupt changes. One important result of this study is the possible relationship between data complexity and forecast accuracy with different models as suggested in the entropy analysis. It is concluded that state space models perform better than long short-term memory networks with highly irregular and more complex surveillance data.

Keywords COVID-19 · State space model · Long short-term memory network · Data complexity

Introduction

The coronavirus disease has had an unprecedented negative impact on the societies and on the economies worldwide in modern times, and the Arab countries were no exception. In fact, currently, it is estimated that these countries may lose around 42 billion dollars and that their unemployment rates could rise by an average of 1.2 percent according to the World Health Organization in their recent report (WHO 2021). But unlike the other North African Arab countries, the Arabian Gulf region has enough resources and wealth to deal with the economic and the social issues triggered by the pandemic. For instance, Saudi Arabia is expected to have a 13 billion dollar plan to support small and medium enterprises. Similarly, the UAE announced a sizeable 27 billion dollar plan to stir their economy and Qatar will put forward a 23 billion dollar package to help their economy private sector. But although the numbers of new cases and deaths have shown a tendency to decline recently in some Gulf countries, most likely because of successful government interventions in providing COVID-19 vaccination earlier that many other countries in the region, the pandemic effects are still being felt by many people and more lives are still being lost as shown in Table 1.

The main question being investigated in this study is related to which methods provide more accurate forecasts of daily coronavirus infections and how do these methods perform when health surveillance records fluctuate considerably. Providing accurate predictions of future trajectories of the spread of the disease could help health officials to implement appropriate, measured, and efficient policies and protocols that may result in reducing the spread of the virus and the number of infections and deaths. This paper makes a contribution to identify the best statistical methods in infectious disease modeling when daily records are subject to abrupt changes. Two statistical methods of forecasting are investigated in this paper. Studies have shown that these models can produce accurate predictions in several epidemiologic applications related to the coronavirus pandemic as shown in Zoabi et al. (2021), Marzouk et al. (2021), Khan et al. (2021), Watson et al. (2021), Ardabili et al. (2020), Sujath et al. (2020). In addition, several statistical and mathematical methods have been suggested in the literature recently to model and to forecast pandemic infections

✉ Sami Khedhiri
samked85@gmail.com

1 School of Mathematical and Computational Sciences,
University of Prince Edward Island, 550 University Avenue,
Charlottetown, PE C1A 4P3, Canada

Table 1 COVID-19 statistics in the Arab Gulf countries (2020/03/24–2021/05/20)

Country	Bahrain	Kuwait	Oman	Qatar	Saudi Arabia	UAE
Infections	209,293	295,861	208,607	214,463	437,569	551,430
Deaths	780	1711	2239	539	7214	1642

and deaths. For instance, Appadu et al. (2021) used Euler' iterative method and cubic spline interpolation to predict the number of COVID-19 infections in South Korea, India, South Africa, Germany, and Italy. Khedhiri (2021) compared the statistical performances of generalized count-data models and zero-inflated models in predicting coronavirus deaths when surveillance data include excess zero counts. Ramazi et al. (2021) designed an adaptive learner based on machine learning techniques and showed that their model produces accurate long-range predictions of COVID-19-related deaths in the US. Rahimi et al. (2021) reviewed some of these methods and their applications, and a recent survey on most popular forecasting models for the coronavirus disease can be found in Shinde et al. (2020).

In this study, we perform an empirical analysis to assess the performance of two alternative methods for short run forecasting of COVID-19 infections in some countries of the Middle East.

The first method is based on a linear state space model where a full description is given about the probabilistic relationship between the observations of a variable of interest and a latent state variable. The second method is related to deep learning and is based on a long short-term memory network. These methods will be presented in the next section.

Methods

Data sources

Data on the cumulative number of infections and deaths related to the pandemic are collected from the Arab countries of the Gulf region. The data are publicly available online and covers the period from March 24, 2020 to May 20, 2021. The countries included in this study are: Bahrain, Kuwait, Qatar, Saudi Arabia, and United Arab Emirates. Therefore, there are 423 observations in total. Next, the daily numbers from the reported cumulative cases are computed for each country.

Statistical models

State space models (SSM) are known to be powerful in solving the problem of learning patterns and predicting behavior in sequential data. Although the linear SSM has been successfully used in various applications in the literature, scholars have recently made significant effort to extend the

method to nonlinear models (Eleftheriadis et al. 2017). Furthermore, with the recent development in precision-based algorithms, methods based on non-linear and non-Gaussian state space models have also been introduced (Chan and Strachan 2012) and for which a new R-package was recently developed (Helske and Vihola 2021) to offer some solutions to the challenging issue of estimation of these models. In addition, Kobayashi (2020) developed a statistical approach based on state space model combined with susceptible–infected–recovered (SIR) model to predict intervention effects for COVID-19 in Japan.

In this study, we consider a Gaussian local level state space linear model which assumes that the observed time series (y_t) is a function of unobservable latent variable (α_t) as defined by model Eqs. (1) and (2):

$$y_t = \alpha_t + \varepsilon_t, \ \varepsilon_t \rightarrow N\left(0, \sigma_\varepsilon^2\right), \tag{1}$$

$$\alpha_{t+1} = \alpha_t + \upsilon_t, \ \upsilon_t \rightarrow N\left(0, \sigma_\upsilon^2\right), \tag{2}$$

where σ_ε^2 and σ_υ^2 are the variances of the observation errors and the state transitions, respectively. This simple model is a special case of a general class of state space models as indicated in Koopman and Durbin (2012). The model is used in this study to forecast the coronavirus infections in the Gulf countries and its prediction performance is assessed. As an alternative method, a long short-term memory network (LSTM) model from deep learning is also considered and its accuracy for forecasting COVID-19 cases is evaluated and compared to SSM. It is common knowledge that LSTM networks are an extension of recurrent neural networks (RNN) which allow to remember and account for past data in memory and to resolve the issue of vanishing gradient that characterizes RNN (Yu et al. 2019). An LSTM model would generally be well suited to fit and predict most time series given time lags of unknown duration. It trains the model using back-propagation and can remember long-term dependence using memory cells and gates. The LSTM can read, write and delete information from its memory which can be thought as a gated cell that decides either to store or to delete information based on its importance. The importance assessment occurs through weights that are learned by the algorithm. There are three gates in the LSTM network: Input, forget and output and they can do back-propagation. The LSTM has the advantage of keeping the training short and the accuracy high. Typically, a long short-term memory model can be represented by Eq. (3) as follows:

$$\text{for}_t = \sigma\left(y_t W_1^{\text{for}} + s_{t-1} W_2^{\text{for}} + b_{\text{for}}\right),$$
$$\text{inp}_t = \sigma\left(y_t W_1^{\text{inp}} + s_{t-1} W_2^{\text{inp}} + b_{\text{inp}}\right),$$
$$\text{out}_t = \sigma\left(y_t W_1^{\text{out}} + s_{t-1} W_2^{\text{out}} + b_{\text{out}}\right),$$
$$M_t = \sigma\left(\text{for}_t * M_{t-1} + \text{inp}_t * H_t\right), \quad (3)$$
$$s_t = \tanh\left(M_t\right) * \text{out}_t,$$
$$H_t = \tanh\left(y_t W_1^{g} + s_{t-1} W_2^{g} + b_H\right),$$

where inp, out, and for denote the input, output and forget gates and y and s refer to the number of input and hidden states, respectively. $W1$ and $W2$ are weight matrices which are adjusted during the network learning phase and b is the bias. H is a candidate hidden state and M is the unit internal memory. Also, tanh and σ are the Tanh layer and the sigmoid layer, respectively, and "*" refers to point wise multiplication of two vectors.

An application of deep learning network method was performed to forecast COVID-19 infections in India (Chandra 2021). Yu et al. (2021) provided a detailed review on recurrent neural networks and LSTM models. More contributions in this field include Yu et al. (2029) for assessing deep learning-based prediction performance of COVID-19 artificial intelligence-based system and Han et al. (2021) for modeling the progression of COVID-19 using Kalman filter method and automated machine learning techniques.

In this paper, the adaptive moment estimation optimizer is applied. This optimizer is an extension of stochastic gradient descendent and is based on adaptive moment estimation and it has two parts: adaptive gradient algorithm and root mean square propagation. An application of LSTM methods with ADAM optimizer to predict wind power generation and temperature with data from Estonia can be found in Misha et al. (2019). The ADAM optimizer computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradient.

Results

The SSM model given by Eqs. (1) and (2) is estimated with KFAS package in *R*. A review of the package and its extensions to the exponential family models is given by Helske (2017). The LSTM model given by Eq. (3) is estimated with *MATLAB* program. Data on COVID-19 daily cases are collected for each country of the Arabian Gulf region. The model estimation is based on sub-sample data which covers the time period from 2020-03-24 to 2021–04-09, thus a total of 381 actual observations. Each initial country data is divided into two parts. The first part (381 observations) is for model estimation, and the second part (remaining 42 observations) is for assessing the forecast performance, then
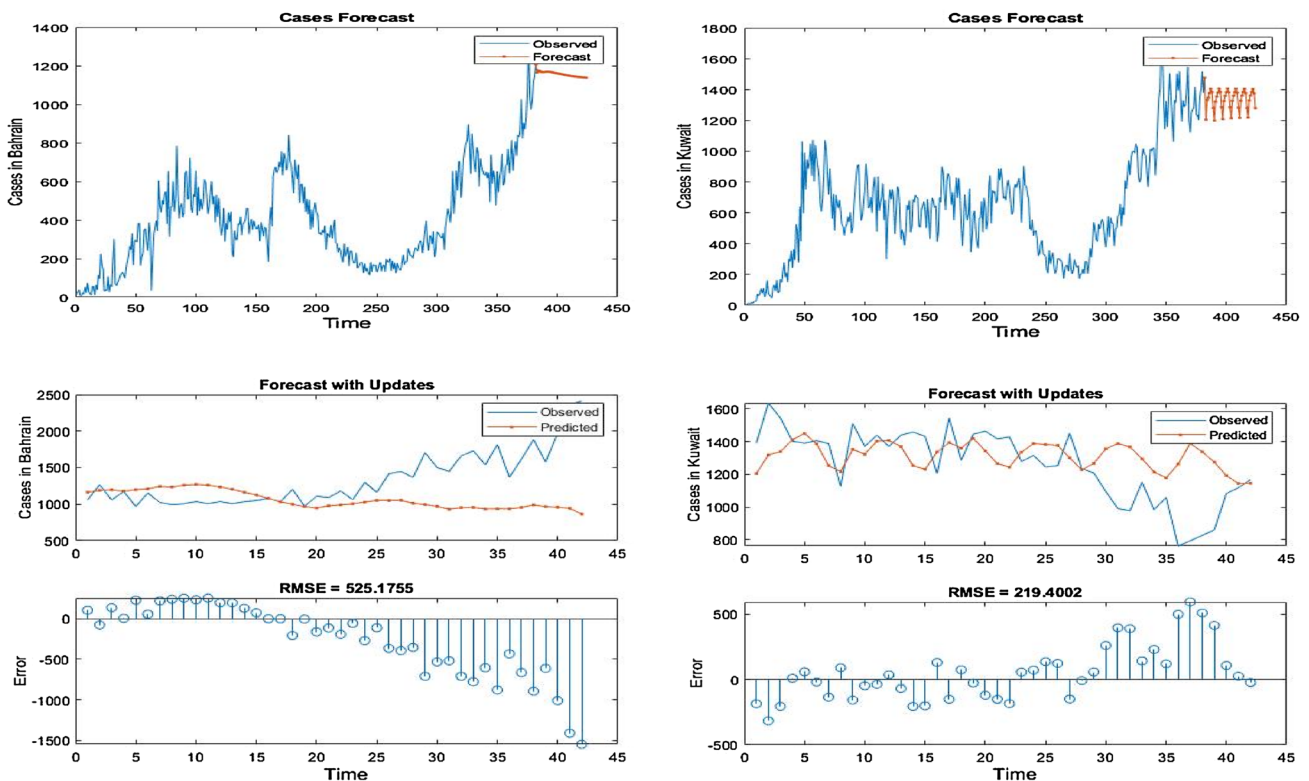

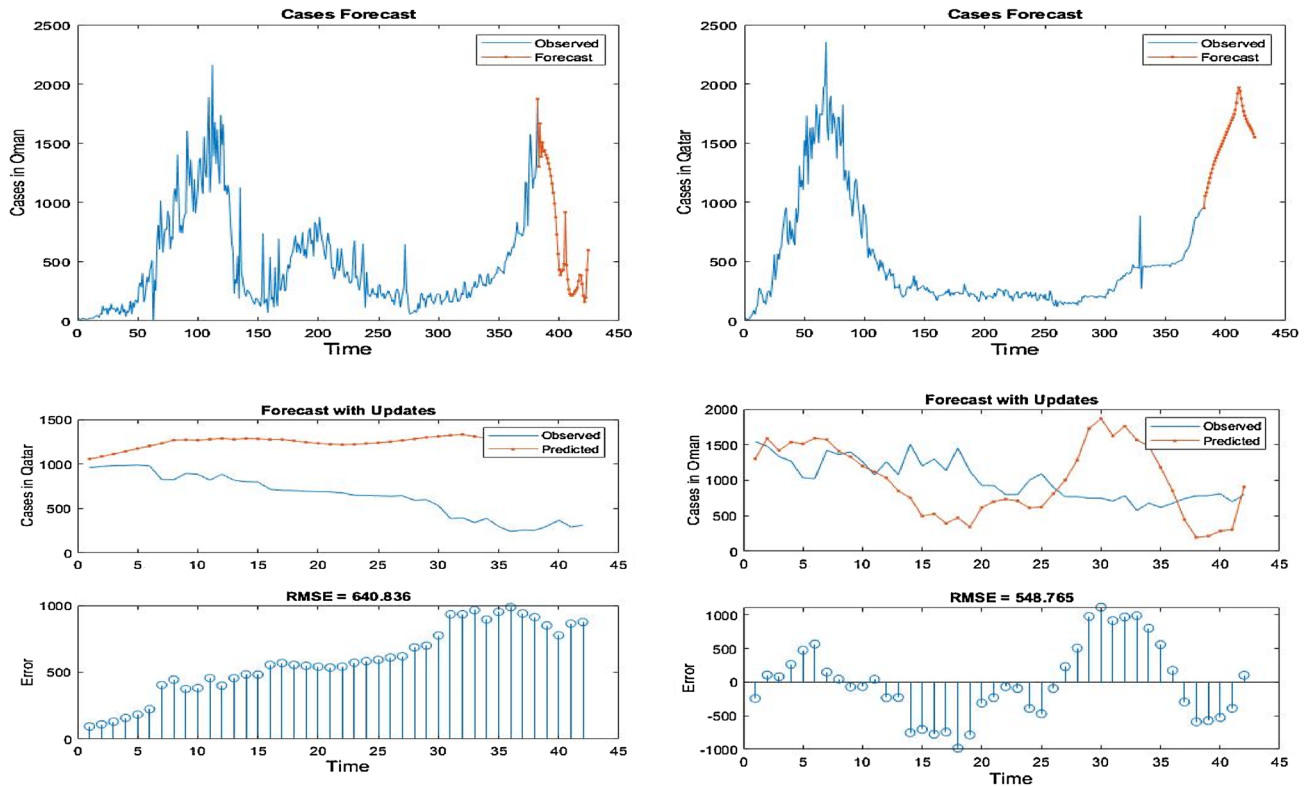**Fig. 1** LSTM COVID-19 forecasts for Bahrain and Kuwait

Fig. 2 LSTM COVID-19 forecasts for Oman and Qatar

the forecasts and the forecast errors are computed from each data. Figures 1, 2, and 3 display the forecasts and the forecast errors of daily cases for each country with the deep learning LSTM method. Figure 4 a, b, c shows infection records and state space model residuals obtained from initial model estimation.

Discussion

It is worth noticing that the fluctuations of daily COVID-19 cases differ significantly between the Gulf countries. This is an important result for the study because the method's forecast accuracy is related to how complex are the data and how important are the fluctuations of infection records. For instance, in the UAE, the daily number of cases at the start of the pandemic was averaging more than 3500 by the end of February 2020, and then dropped by nearly a half in mid-April of the same year. The number continues to decline steadily to reach just about 1000 cases a day by May 2021. For Saudi Arabia, the daily average was around 5000 in mid-June 2020 and then it started to decline almost steadily to around just 1000 cases in May 2021. A similar situation can describe the pattern of daily cases in Kuwait. After fluctuating around 700 cases a day from May to August 2020,

the number of COVID-19 infections decreased and then picked up in April 2021 to reach a staggering average of more than 1500 cases a day. However, this was followed by a sharp decrease in May 2021. The common feature for these three countries is that their daily average numbers of cases tend to decline in the last month of the sample data. But the opposite pattern characterizes the remaining three Gulf countries, for which the reported coronavirus statistics show a different pattern. In Bahrain, the average number of daily cases was just about 500 before summer 2020, then it jumped to nearly the double of that number in March and April of 2021. Unlike Saudi Arabia, for the UEA and Kuwait, the number of daily COVID-19 cases continued to increase sharply in May 2021. The situation in Oman was not very different although to a slightly lesser degree than in Bahrain. After a sharp and steady increase in daily cases to reach over 2000 infections, the numbers declined in summer 2020, but then increased again sharply in April 2021, despite a sign of small decrease in May 2021. In Qatar, the start of the pandemic is marked by a sharp increase in the number of daily COVID-19 infections to reach nearly 2000 cases a day in May 2020. The numbers, however, showed a sign of decline in August 2020. But similarly to Bahrain, the number of cases jumped again in April 2021. However, recent statistics show that the daily average of infections in Qatar
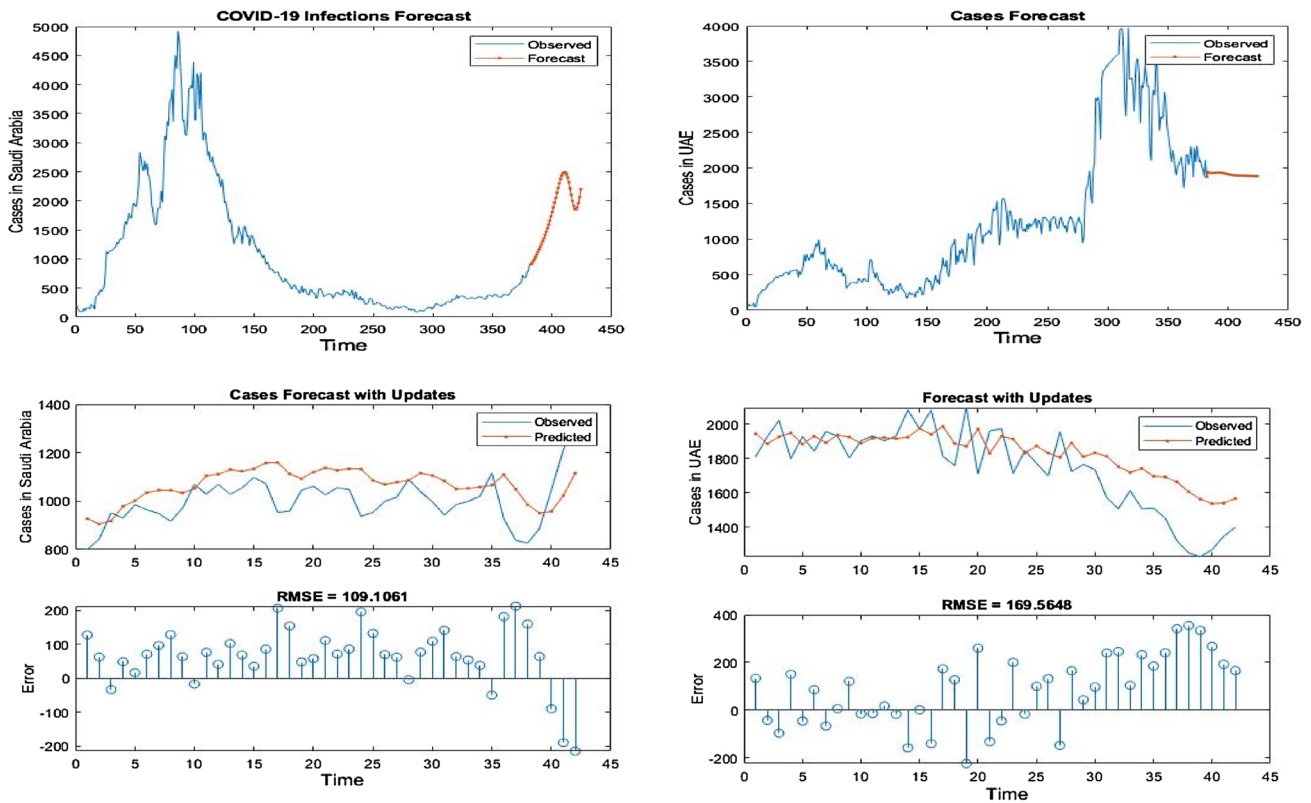
Fig. 3 LSTM COVID-19 forecasts for Saudi Arabia and the UAE

has slightly deceased in May 2021. Table 2 displays the root mean square forecast errors computed for each country's infection records and based on forecast errors obtained from within sample forecasts.

Although the forecast performance of the two methods is shown to be comparable for Saudi Arabia and the UEA, it is clear that SSM outperforms LSTM forecasts for the other countries. Furthermore, the long short-term memory model gave a remarkable poor performance particularly for three countries; namely Bahrain, Oman and Qatar, for which the values of LSTM root mean square forecast errors, are more than three times higher than those of the state space model forecasts. The left panel of Fig. 1 displays COVID-19 infections data and forecasts with LSTM model for Bahrain. It clearly indicates a widening gap between model predictions and actual data in particular during the second half of the data sample which is characterized by a steep decrease followed by a sharp increase in the number of infections. The number of daily infections in Kuwait and their forecasts with LSTM are depicted in the right panel of Fig. 1 and it shows a smaller drift between actual data and predicted values compared to Bahrain. In fact, both curves close in on each other towards the end of the sample. This is different from LSTM forecast of the COVID-19 infections in Oman, where the LSTM deep learning network predicts an overall decreasing

pattern which is not supported by the actual data as shown in the left panel of Fig. 2. Similarly, forecast values and actual number of infections move in opposite directions for Qatar, except that in this case the LSTM model predicts wrongfully an increasing pattern that is also not supported by the data as illustrated in the right panel of Fig. 2. For Saudi Arabia and the UAE, the LSTM forecast accuracy is better than the results obtained for the other Gulf countries. In fact, Fig. 3 displays forecast error curves not drifting too far from zero for Saudi Arabia and the UAE.

The forecast and the fit performances of the SSM model are described in Fig. 4 a, b, c. The curves in black refer to infection records and the curves in red are for model residuals. Except for the case of infections in Qatar for which Fig. 4 a,b, c show a noticeable abrupt change in the model residuals for a number of daily infections at the beginning of the third quarter of the sample, the results depict generally smooth error curves for the other countries.

In an effort to understand what could possibly explain this large gap in the prediction accuracy between the two methods, an entropy analysis to evaluate the complexity of each COVID-19 daily infection data is investigated in this study. Our objective is to verify if there is a relationship between forecast performance and data complexity.
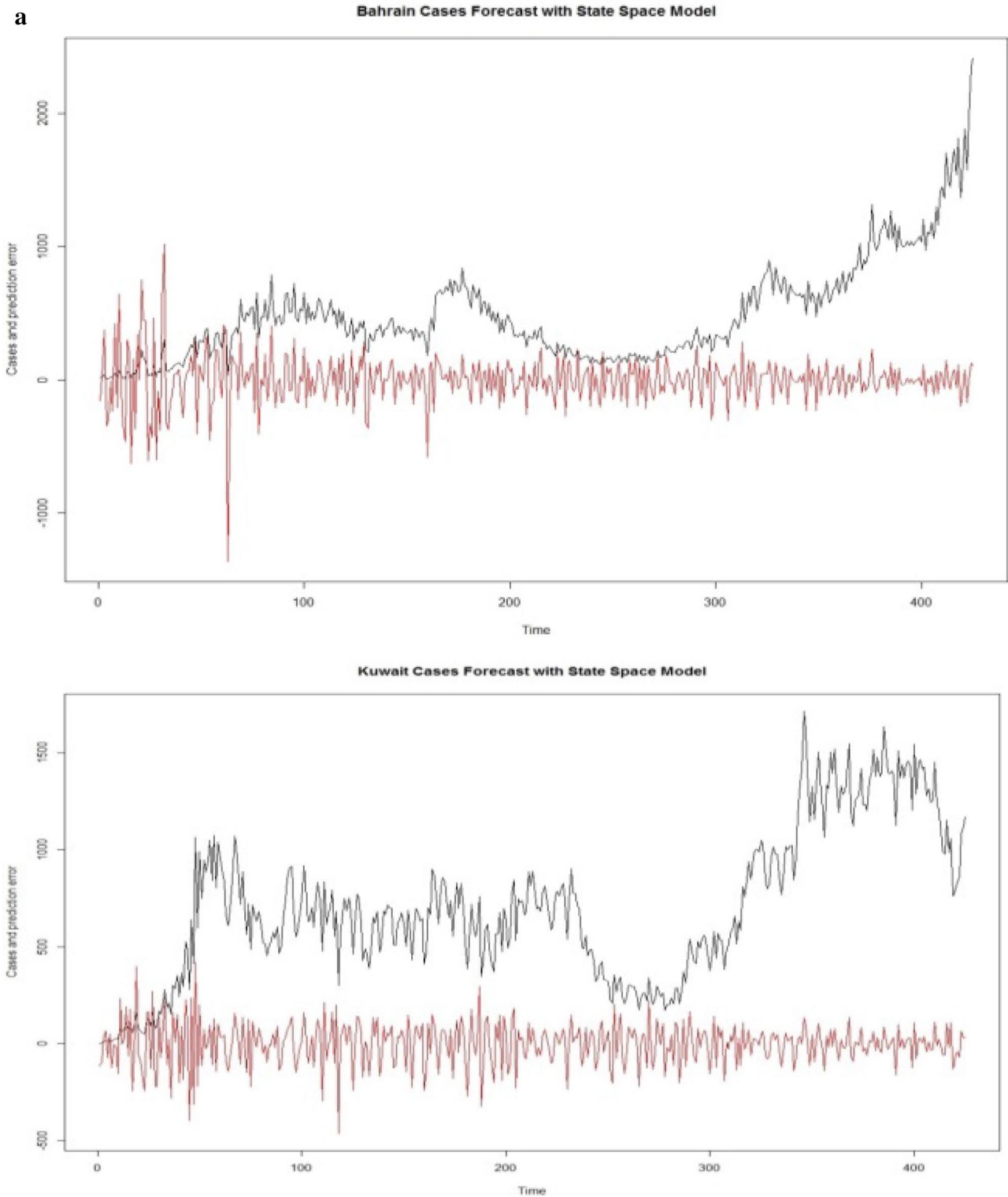
a

Bahrain Cases Forecast with State Space Model

Kuwait Cases Forecast with State Space Model

Fig. 4 SSM COVID-19 forecasts for the Arabian Gulf countries. **a** SSM model estimation of COVID-19 infections for Bahrain and Kuwait, **b** SSM model estimation of COVID-19 infections for Oman and Saudi Arabia, and **c** SSM model estimation of COVID-19 infections for Qatar and the UAE
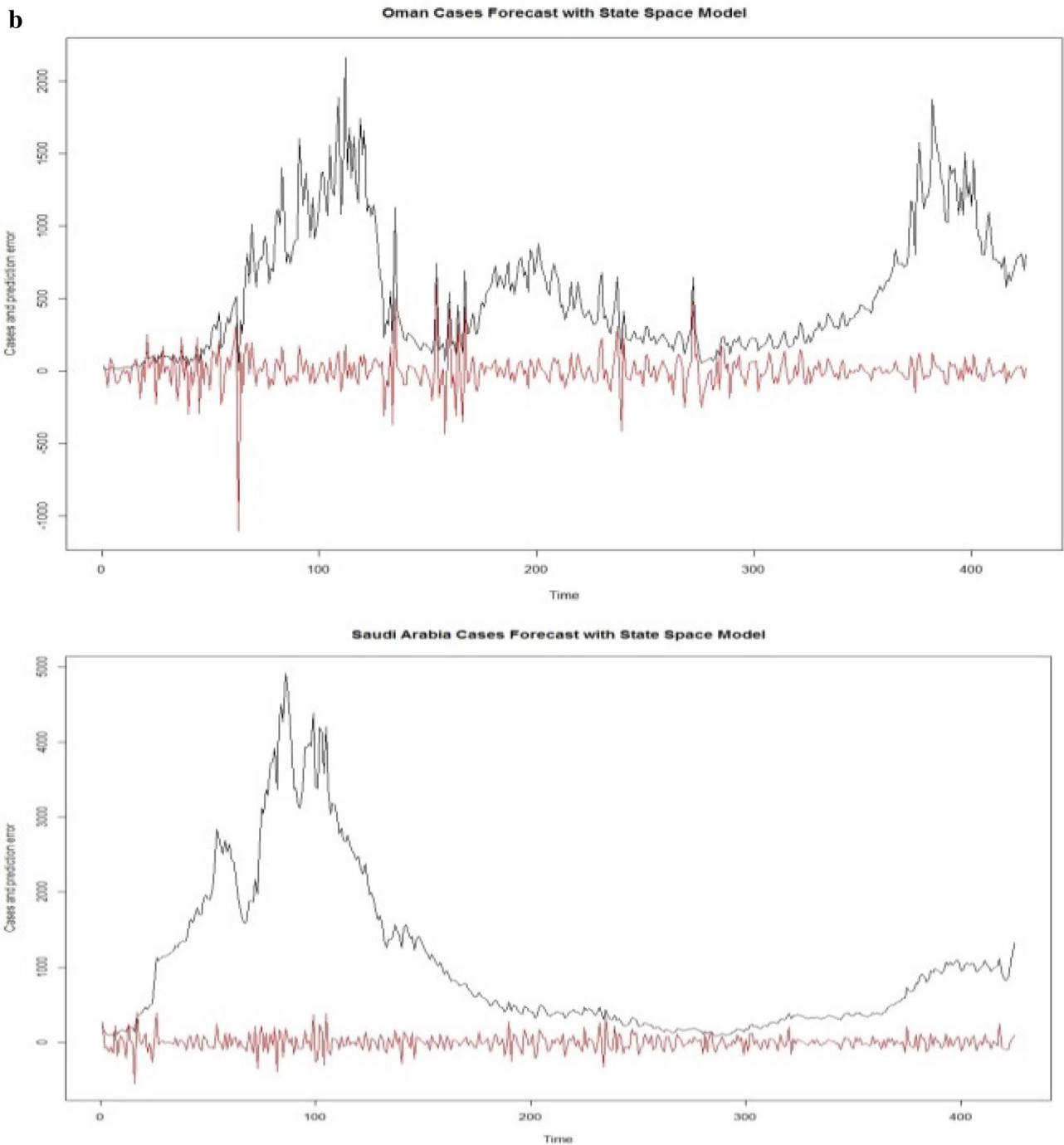
b

Oman Cases Forecast with State Space Model

Saudi Arabia Cases Forecast with State Space Model

Fig. 4 (continued)

Measuring complexity of a time series can lead to crucial insights into the functioning of the system under investigation as indicated in Nagaraj et al. (2013). There are several measures of complexity that have been suggested in the literature. These include approximate entropy, sample entropy, fuzzy entropy and permutation entropy. It was found that these measures were less ambitious but more practical alternatives to the classical techniques for the analysis of nonlinear dynamical systems, like correlation dimension, Lyapunov exponents, and nonlinear prediction methods (Xiong et al. 2007). In addition, researchers suggested that the popularity of entropy measures stems from their applicability to short and noisy processes with stochastic components such as those describing the dynamical activity
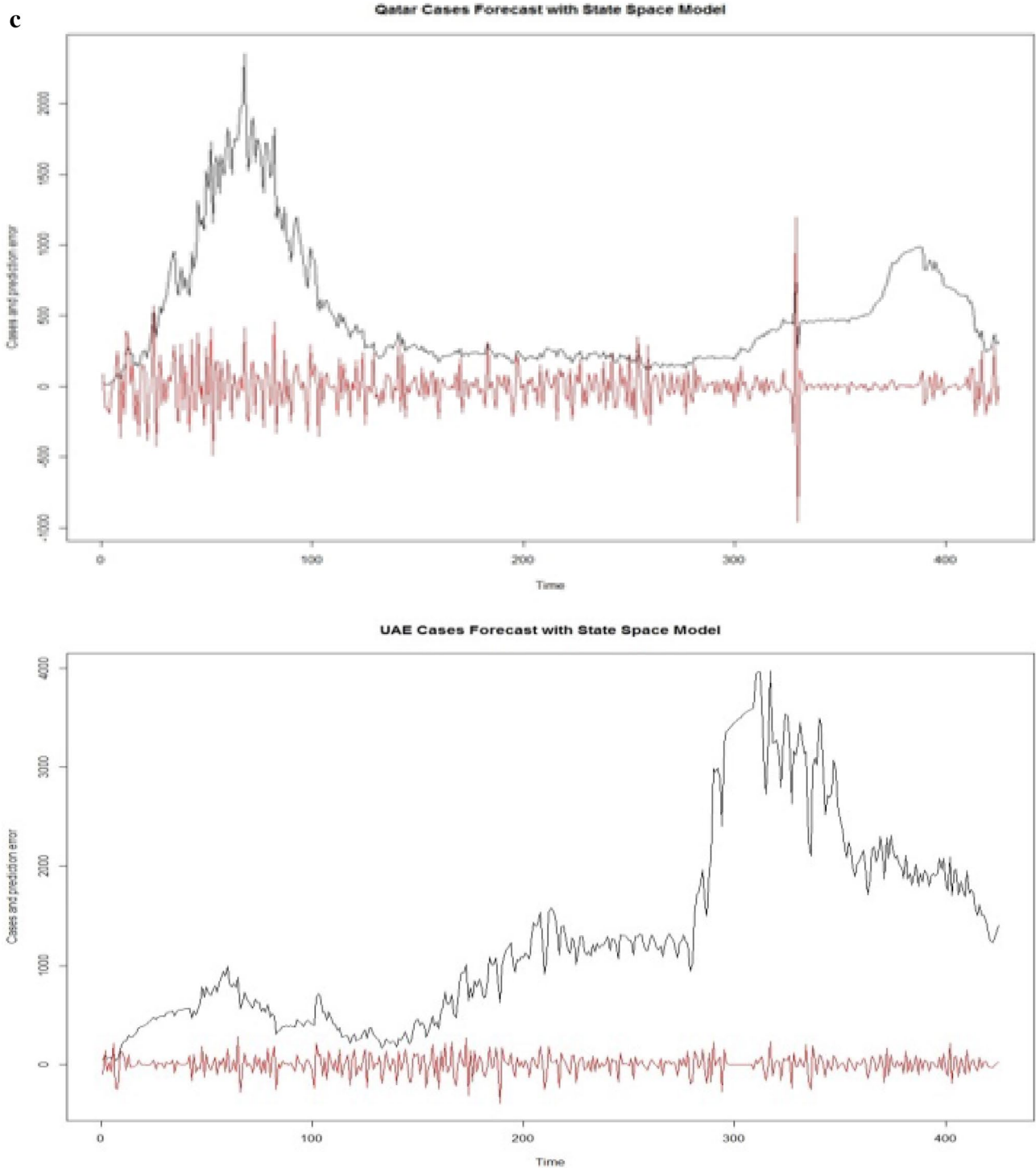
c

Qatar Cases Forecast with State Space Model

UAE Cases Forecast with State Space Model

Fig. 4 (continued)

of real-world system. Sample entropy, in particular, is still considered as one of the most powerful tools for analyzing the complexity and irregularity in various data applications (Chen et al. 2019).

Some extensions of sample entropy have also been introduced in the literature with an interesting application of multi-scale entropy to study climate data (Balzter et al. 2015). Furthermore, flexible multi-scale entropy for sensor networks based on a novel similarity function was also

Table 2 Root mean square error (RMSE): assessment of forecast performance of alternative prediction methods for COVID-19 infections in the Arab Gulf countries

Country	LSTM (RMSFE)	SS Model (RMSFE)	Sample entropy estimates
Bahrain	525.18	184.68	0.28477
Kuwait	219.40	112.29	0.24762
Oman	548.77	123.80	0.30338
Qatar	640.84	156.16	0.34898
Saudi Arabia	109.11	111.19	0.06725
UAE	169.57	101.06	0.14732

developed (Zhou et al. 2017). In this paper, a sample entropy approach is applied to estimate the randomness and to study the complexity of COVID-19 public health surveillance data in the Arab countries without any previous knowledge about the source generating the dataset.

Let $\{X_T\}$ be a times series given by $X_T = \{x_1, x_2, ..., x_T\}$, and suppose we select 2 sequences of n consecutive observations, $X_n(t) = \{x_t, x_{t+1}, ..., x_{t+n-1}\}$ and $X_n(\tau) = \{x_\tau, x_{\tau+1}, ..., x_{\tau+n-1}\}$, where t and τ are in $[1, ..., T-n]$ with $t \neq \tau$. Next, the maximum distance for each sequence is computed and is compared to some tolerance level λ for repeated sequences counting. The counts for $X_n(t)$ are denoted by $C_t^n(\lambda)$ and the maximum distance is given by,

$$\text{distance}\big(X_n(t), X_n(\tau)\big) = \text{maximum}[|x_{t+i}, x_{\tau+i}|]$$
$$\leq \lambda(0 \leq i \leq n-1, \text{ for } \lambda \geq 0).$$

We choose the tolerance level λ to be approximated by 0.15 times the standard deviation of the original time series $\{X_T\}$, following Richman and Moorman (2000) and Hansen et al. (2017) who also extended their simulation analysis to study single variable and multivariate multi-scale entropy.

Let $C^n(\lambda)$ be the average amount of $C_t^n(\lambda)$ and let $C^{n+1}(\lambda)$ be the average of $(n+1)$ consecutive observations. The sample entropy, SE, is then defined as follows:

$$\text{SE}(T, n, \lambda) = -\ln\left[\frac{C^{n+1}(\lambda)}{C^n(\lambda)}\right] = -\ln\left[\frac{(T-n)\sum_{t=1}^{T-n-1}C_t^{n+1}(\lambda)}{(T-n-1)\sum_{t=1}^{T-n}C_t^n(\lambda)}\right].$$

The parameter n refers to the length of repeated mode in the series and the tolerance parameter λ sets the limitation condition of repeated mode.

The last column of Table 2 shows the estimated values of sample entropy for each country data. The computations were performed in R. It is shown that the higher is the complexity measure, the larger is the gap in root mean square error between the two methods. In addition, our findings

point to a poor prediction performance when the data exhibit high randomness and abrupt changes in the daily infections. Therefore, we can determine two clusters of countries for the group of six Arabian Gulf countries. The first cluster which is characterized by more complex data includes Bahrain, Oman and Qatar and for which LSTM neural network was unable to provide accurate infection predictions. The second cluster includes Saudi Arabia, UAE, and perhaps to a lesser degree Kuwait. These countries have fewer high fluctuations in their daily cases data and for which the results show the forecasting performance of deep learning LSTM model and state space model were fairly close. We recognize the limitation of our approach and we can state that further complexity analysis, for instance including a multi-scale entropy approach, could perhaps provide more evidence on the relationship between the difference in forecast performance and data irregularity. We confine our study to sample entropy analysis, and for future research directions more comprehensive complexity analysis of public health surveillance may be considered to further investigate the issue of data irregularity and model forecast accuracy.

Conclusion

The coronavirus pandemic has had a notable negative effect on humankind and has claimed the lives of many people. Numerous research papers and studies have emerged recently in an effort to assess the economic and social impacts of this deadly disease and to predict accurately its future course to implement health policies and protocols accordingly. However, statistical methods may differ significantly in their forecasting performance due to the high fluctuations of daily infection records. Thus, it is crucial to identify which methods are more accurate than others. This paper makes a contribution by conducting a statistical analysis to evaluate the prediction performance for two of these methods with COVID-19 data from the Arabian Gulf countries. The first method is based on a local level state space model assuming Gaussian probability distribution. The second is a deep learning method and is based on long short-term memory neural networking. For each model, short-term infection forecasts are determined and root mean square errors are computed. The results show that state space models produce smaller root mean square errors and, thus, outperform the long short-term memory networks. One explanation is the inability of LSTM to predict accurately in the presence of highly complex health surveillance data.

Declarations

Conflict of interest The author has no conflict of interest for this research.

Ethics approval This research uses publicly available data and thus, ethics approval is not required.

References

Appadu AR, Kelil AS, Tijani YO (2021) Comparison of some forecasting methods for COVID-19. Alex Eng J 60(1):1565–1589. https://doi.org/10.1016/j.aej.2020.11.011

Ardabili S, Mosavi A, Ghamisi P, Ferdinand F et al (2020) COVID-19 outbreak prediction with machine learning. Algorithms 13:249. https://doi.org/10.3390/a13100249

Balzter H, Tate N, Kaduk J et al (2015) Multi-scale entropy analysis as a method for time series analysis of climate data. Climate 3(1):227–240. https://doi.org/10.3390/cli3010227

Chan J, Strachan R (2012) Estimation in non-linear non-Gaussian state space models with precision-based models. Munich Personal RePEc Archive, MPRA_paper_39360.pdf. https://mpra.ub.uni-muenchen.de/39360/

Chandra R, Jain A, Chauhan DS (2021) Deep learning via LSTM models for COVID-19 infection forecasting in India. https://arxiv.org/pdf/2101.11881.pdf

Chen C, Sun S, Cao Z et al (2019) A comprehensive comparison and overview of R packages for calculating sample entropy. Biol Methods Protocols. https://doi.org/10.1093/biomethods/bpz016

Eleftheriadis S, Nicholson T, Deisenroth M, et al (2017) Identification of Gaussian process state space models. 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, USA. https://papers.nips.cc/paper/2017/file/1006ff12c465532f8c574aeaa4461b16-Paper.pdf

Han T, Gois FNB, Oliveira R et al (2021) Modeling the progression of COVID-19 deaths using Kalman filter and autoML. Soft Comput. https://doi.org/10.1007/s00500-020-05503-5

Hansen C, Wei Q, Shieh JS et al (2017) Sample entropy, univariate and multivariate multiscale entropy in comparison with classical postural sway parameters in young healthy adults. Front Hum Neurosci 11:206. https://doi.org/10.3389/fnhum.2017.00206

Helske J (2017) KFAS: exponential family state space models in R. J Stat Soft. https://doi.org/10.18637/jss.v078.i10

Helske J, Vihola M (2021) Bssm: Bayesian inference of non-linear and non-Gaussian state space models in R. R J. https://arXiv.org/abs/2101.08492v2/

Khan F, Ali S, Saeed A, Kumar R, Khan AW (2021) Forecasting daily new infections, deaths and recovery cases due to COVID-19 in Pakistan by using Bayesian dynamic linear models. PLoS ONE 16(6):e0253367. https://doi.org/10.1371/journal.pone.0253367

Khedhiri S (2021) Statistical modeling of COVID-19 deaths with excess zero counts. Epidemiol Methods. https://doi.org/10.1515/em-2021-0007

Kobayashi G, Sugasawa S, Tamae H et al (2020) Predicting intervention effects for COVID-19 in Japan: a state space modeling approach. Biosci Trends 14(3):174–181. https://doi.org/10.5582/bst.2020.03133

Koopman SJ, Durbin J (2012) Time series analysis by state space methods. Second edition. Oxford University Press

Marzouk M, Elshaboury N, Abdel-Latif M, Azab S (2021) Deep learning model for forecasting COVID-19 outbreak in Egypt. Process Saf Environ Prot 153:363–375. https://doi.org/10.1016/j.psep.2021.07.034

Misha S, Bordin C, Taharaguchi K et al (2019) Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature. Energy Rep 6(3):273–286

Nagaraj N, Balasubramanian K, Dey S (2013) A new complexity measure for time series analysis and classification. Eur Phys J Spec Top 222:847–860. https://doi.org/10.1140/epjst/e2013-01888-9

Painuli D, Mishra D, Bhardwaj S, Aggarwal M (2020) Forecast and prediction of COVID-19 using machine learning. Data Sci COVID-19. https://doi.org/10.1016/B978-0-12-824536-1.00027-7

Rahimi I, Chen F, Gandomi AH (2021) A review on COVID-19 forecasting methods. Neural Comput Appl. https://doi.org/10.1007/s00521-020-05626-8

Ramazi P, Haratian A, Meghdadi M et al (2021) Accurate long-range forecasting of COVID-19 mortality in the USA. Sci Rep. https://doi.org/10.1038/s41598-021-91365-2

Richman J, Moorman J (2000) Physiological time series analysis using approximate entropy and sample entropy. Am J Physiol-Heart Circ Physiol 278(6):2039–2049. https://doi.org/10.1152/ajpheart.2000.278.6.H2039

Shinde G, Kalamkar A, Mahalle P et al (2020) Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. SN Comput Sci. https://doi.org/10.1007/s42979-020-00209-9

Sujath R, Chatterjee JM, Hassanien AE (2020) A machine learning forecasting model for COVID-19 in India. Stoch Environ Res Risk Assess 34:959–972. https://doi.org/10.1007/s00477-020-01827-8

Watson GL, Xiong D, Zhang L, Zoller JA, Shamshoian J, Sundin P et al (2021) Pandemic velocity: forecasting COVID-19 in the US with a machine learning & Bayesian time series compartmental model. PLoS Comput Biol 17(3):e1008837. https://doi.org/10.1371/journal.pcbi.1008837

WHO's Situation Report of 25 May (2021) Arab Countries Respond to COVID-19. https://www.arabstates.undp.org/content/rbas/en/home/coronavirus.html

Xiong W, Faes L, Ivanov P (2017) Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: effects of artifacts, non-stationarity, and long range correlations. Phys Rev E. https://doi.org/10.1103/PhysRevE.95.062114

Yu Y, Si X, Hu C et al (2019) A review of recurrent neural networks: LSTM cells and network. Neural Comput 31(7):1235–1270. https://doi.org/10.1162/neco_a_01199

Yu CS, Chang SS, Chang TH et al (2021) A COVID-19 pandemic artificial intelligence–based system with deep learning forecasting and automatic statistical data acquisition: development and implementation study. J Med Internet Res 23(5):e27806. https://doi.org/10.2196/27806

Zhou R, Yang C, Wan J et al (2017) Measuring complexity and predictability of time series with flexible multiscale entropy for sensor networks. Sensors. https://doi.org/10.3390/s17040787

Zoabi Y, Deri-Rozov S, Shomron N (2021) Machine learning-based prediction of COVID-19 diagnosis based on symptoms. Npj Digit Med. https://doi.org/10.1038/s41746-020-00372-6

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.