Infectious Disease Modelling 9 (2024) 816–827

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Predicting influenza-like illness trends based on sentinel surveillance data in China from 2011 to 2019: A modelling and comparative study[1]

Xingxing Zhang [a, b, c, d, 1], Liuyang Yang [a, b, c, e, 1], Teng Chen [f, 1], Qing Wang [a, b, c], Jin Yang [a, b, c], Ting Zhang [a, b, c], Jiao Yang [a, b, c], Hongqing Zhao [d], Shengjie Lai [g], Luzhao Feng [a, b, c, *], Weizhong Yang [a, b, c, **]

[a] School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, 100073, China
[b] State Key Laboratory of Respiratory Health and Multimorbidity, China
[c] Key Laboratory of Pathogen Infection Prevention and Control (Peking Union Medical College), Ministry of Education, China
[d] National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, China
[e] Department of Management Science and Information System, Faculty of Management and Economics, Kunming University of Science and Technology, Kunming, 650506, China
[f] Department of Applied Mathematics & Statistics, Stony Brook University, Stony Brook, NY, 11794-3600, USA
[g] WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, SO17 1BJ, UK

ARTICLE INFO

Article history:
Received 20 November 2023
Received in revised form 25 April 2024
Accepted 26 April 2024
Handling editor: Yijun Lou

Keywords:
Influenza-like illness
Influenza
Sentinel surveillance
China
Predicting
Modeling

ABSTRACT

Background: Influenza is an acute respiratory infectious disease with a significant global disease burden. Additionally, the coronavirus disease 2019 pandemic and its related non-pharmaceutical interventions (NPIs) have introduced uncertainty to the spread of influenza. However, comparative studies on the performance of innovative models and approaches used for influenza prediction are limited. Therefore, this study aimed to predict the trend of influenza-like illness (ILI) in settings with diverse climate characteristics in China based on sentinel surveillance data using three approaches and evaluate and compare their predictive performance.

Methods: The generalized additive model (GAM), deep learning hybrid model based on Gate Recurrent Unit (GRU), and autoregressive moving average-generalized autoregressive conditional heteroscedasticity (ARMA—GARCH) model were established to predict the trends of ILI 1-, 2-, 3-, and 4-week-ahead in Beijing, Tianjin, Shanxi, Hubei, Chongqing, Guangdong, Hainan, and the Hong Kong Special Administrative Region in China, based on sentinel surveillance data from 2011 to 2019. Three relevant metrics, namely, Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and R squared, were calculated to evaluate and compare the goodness of fit and robustness of the three models.

Results: Considering the MAPE, RMSE, and R squared values, the ARMA—GARCH model performed best, while the GRU-based deep learning hybrid model exhibited moderate

* Corresponding author. School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, 100073, China.
** Corresponding author. School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, 100073, China.
E-mail addresses: fengluzhao@cams.cn (L. Feng), yangweizhong@cams.cn (W. Yang).
Peer review under responsibility of KeAi Communications Co., Ltd.
[1] These authors contributed equally to the study.

https://doi.org/10.1016/j.idm.2024.04.010

performance and GAM made predictions with the least accuracy in the eight settings in China. Additionally, the models' predictive performance declined as the weeks ahead increased. Furthermore, blocked cross-validation indicated that all models were robust to changes in data and had low risks of overfitting.

Conclusions: Our study suggested that the ARMA—GARCH model exhibited the best accuracy in predicting ILI trends in China compared to the GAM and GRU-based deep learning hybrid model. Therefore, in the future, the ARMA—GARCH model may be used to predict ILI trends in public health practice across diverse climatic zones, thereby contributing to influenza control and prevention efforts.

Abbreviations

(NPIs) non-pharmaceutical interventions
(ILI) influenza-like illness
(GAM) generalized additive model
(GRU) Gate Recurrent Unit
(ARMA—GARCH) autoregressive moving average-generalized autoregressive conditional heteroscedasticity
(MAPE) Mean Absolute Percentage Error
(RMSE) Root Mean Squared Error
(COVID-19) coronavirus disease 2019
(WHO) World Health Organization
(RT-PCR) Real-time reverse transcription PCR
(SAR) Special Administrative Region
(CI) confidence interval

1. Introduction

Influenza is an acute respiratory infectious disease caused by influenza viruses that circulate among people worldwide (Uyeki et al., 2022). This disease is characterized by its robust infectivity, rapid transmission, and antigen variation. Seasonal influenza reportedly occurs with an annual attack rate of 5–10% in adults and 20–30% in children globally (World Health Organization). Annual influenza epidemics commonly occur during the cold winter in temperate climates worldwide. Additionally, year-round influenza activity can be observed in tropical and subtropical areas, peaking at different times. Influenza was the first infectious disease, systematically monitored globally (World Health Organization). Influenza-like illness (ILI), which is characterized by patients presenting with fever of $\geq$38 °C and cough or sore throat (World Health Organization), is a proxy indicator of influenza activity. Therefore, a sentinel ILI surveillance system was established to collect epidemiological data on influenza and monitor the influenza trends.

The coronavirus disease 2019 (COVID-19) pandemic and its related non-pharmaceutical interventions (NPIs) have altered the predictable circulation pattern of influenza globally. Notably, influenza activity was significantly reduced in northern and southern China and the United States during the 2019–2020 season (Feng et al., 2021). In the World Health Organization (WHO) European region, the 2019–2020 influenza season was prematurely concluded by the COVID-19 pandemic and was characterized by low-level influenza activity (Adlhoch et al., 2021). Furthermore, low exposure to the influenza virus and waning immunity over time add uncertainty to the future influenza circulation. For instance, the influenza peak season was significantly earlier in the 2022–2023 season than in the previous five seasons and is associated with a higher hospitalization rate in the United States (Centers for Disease Control and Prevention, 2023). Although COVID-19 remains a persistent global health threat (Lazarus et al., 2022), the rebound and resurgence of influenza are emerging or imminent. Therefore, predicting influenza or ILI trends with optimal models and techniques using surveillance data is crucial to inform the early warning and intervention strategies (Hay & McCauley, 2018) for influenza, particularly in the pandemic or post-pandemic era.

Numerous innovative models and approaches have been applied to influenza prediction. For example, the deep learning hybrid model (Reichstein et al., 2019)—a machine learning algorithms—has increasingly become popular because of its intelligent learning ability (He et al., 2022; Tang Chaofan, Yang, Tang, & Zhao, 2023). Meanwhile, the autoregressive moving average-generalized autoregressive conditional heteroscedasticity (ARMA—GARCH) model (Bollerslev, 1986 combines two classic time-series models and is widely used for fitting data with periodic fluctuations in financial and economic fields. The generalized additive model (GAM) can adapt to complex nonlinear relationships and make superior predictions capable of

incorporating multiple influencing factors (Iva et al., 2018). These three models can effectively predict time series data with periodic fluctuations and perform well in different research fields. However, studies on the performance of these models in predicting the ILI trend are limited.

Therefore, this study is aimed to predict the ILI trend using sentinel surveillance data in settings with diverse climate characteristics in China with three approaches: the generalized additive model (GAM), deep learning model based on Gate Recurrent Unit (GRU), and ARMA-GARCH model. Additionally, the predictive performance of these models was assessed. Our findings will add evidence to the innovation and application of ILI- and influenza-predicting techniques while also helping to addressthe constant threat of emerging infectious diseases, such as COVID-19 and other infectious diseases.

2. Materials and methods

2.1. Data collection

2.1.1. Influenza surveillance dataset

ILI and virological surveillance data in China were obtained from the National Influenza Surveillance Network from 2011 to 2019. The National Influenza Surveillance Network in mainland China, led by China Center for Disease Control and Prevention (China CDC, Beijing), comprises 554 sentinel hospitals and 407 network laboratories (National Health Commission of the People's Republic of China) (Yu et al., 2013). Influenza activity levels and trends are monitored using ILI data from sentinel hospitals and virological data from the Influenza Network Laboratory (Feng et al., 2021).

The weekly proportions of ILI cases in the Hong Kong Special Administrative Region (SAR) were obtained from the Centre for Health Protection of the Hong Kong SAR, based on sentinel surveillance conducted at the General Outpatient Clinics/ Private Medical Practitioner Clinics (Center for Health Protection/Department of Health of the Government of the Hong Kong Special Administrative Region; Lau et al., 2008). The influenza sentinel surveillance network comprised approximately 60 outpatient clinics (Cowling et al., 2013; 2020). Each week, data on the proportions of ILI cases per 1000 outpatient consultations (Wu et al., 2017))and influenza test positivity specimens by type, subtype and specimens tested (Yang et al., 2015) are reported.

As the influenza seasonality patterns in different climates were heterogeneous, seven provinces/megacities in mainland China and the Hong Kong SAR from three different climatic zones were selected as the study settings. Among them, Beijing, Tianjin, and Shanxi are located in the temperate region, Hubei, Chongqing, most of Guangdong, and the Hong Kong SAR are located in the subtropical region, and Hainan is in the tropical region (Fig. 1). In mainland China, each surveillance year comprises a 12-month interval as follows: from week 14 of one year to week 13 of the following year. For consistency within this study, the surveillance year in the Hong Kong SAR was similar to that in mainland China.

2.1.2. Meteorological dataset

Meteorological data, including mean temperature, absolute humidity, relative humidity, and other meteorological factors in the eight provinces/megacities between 2011 and 2019 were obtained from the China Meteorological Data Service Center (China Meteorological Administration) and the National Centers for Environmental Information (National Oceanic and Atmospheric Administration). Temperature and humidity play a more important role in influenza transmission (Shaman & Kohn, 2009). Thus, before modeling, collinearity between mean temperature, relative humidity, and absolute humidity was analyzed using the Pearson correlation test (Tables S1–S8). When the correlation coefficient between a pair of variables was >0.7, only one of them was included in the model. Therefore, absolute humidity was incorporated into the GAM models in the eight settings.

2.1.3. Public and school holiday dataset

The holiday data in the eight provinces/megacities between 2011 and 2019 were obtained from the WorldPop Data repository (WorldPop, 2022) and a relevant previous study (Lai et al., 2022). These datasets were established as a time series to record the number of days each week containing public or school holidays.

2.1.4. Population density dataset

The yearly population density data for the eight provinces/megacities between 2011 and 2019 was downloaded from the statistical yearbooks released by the Statistics Bureau of Beijing (Beijing Municipal Bureau Statistics), Tianjin (Tianjin Municipal Bureau Statistics), Shanxi (Shanxi Provincial Bureau Statistics), Hubei (Hubei Municipal Bureau Statistics), Chongqing (Chongqing Provincial Bureau Statistics), Guangdong (Guangdong Provincial Bureau Statistics), Hainan (Hainan Municipal Bureau Statistics), and Statistics Department of the Hong Kong SAR (Census and Statistics, Department of the Government of the Hong Kong SAR).

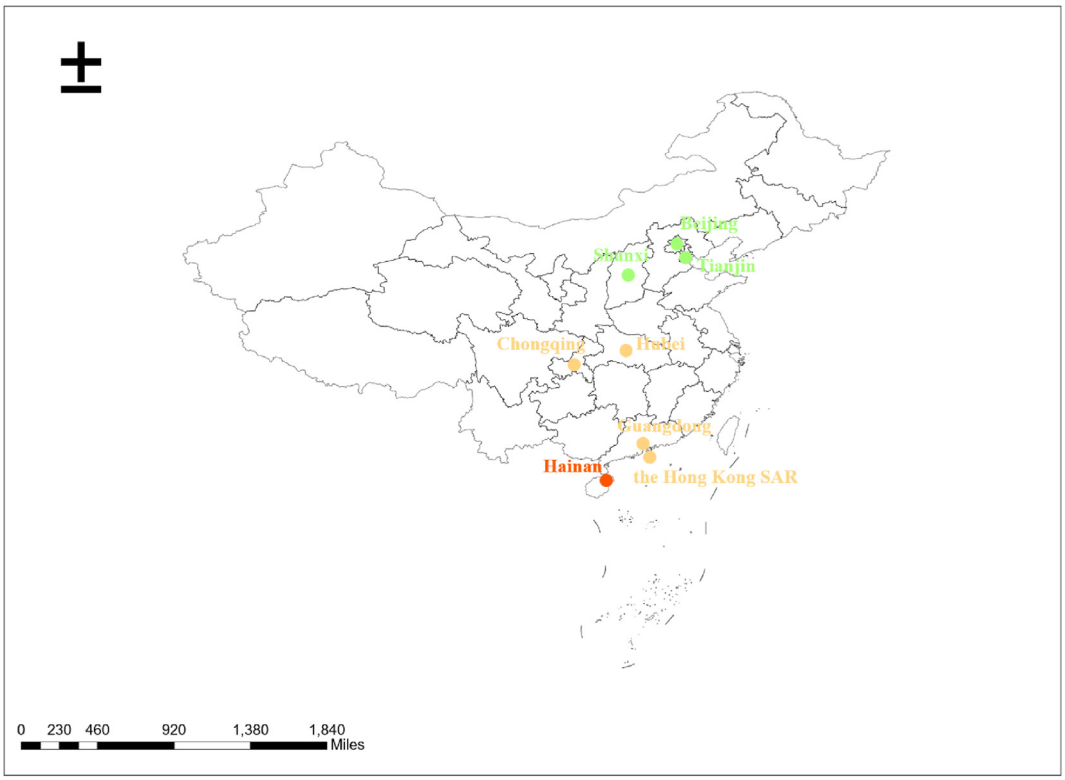

Fig. 1. Map indicating the geographical locations of the eight provinces/megacities in China in the study. The colors illustrate the following climate zones: temperature (green), subtropical (yellow), and tropical (red).

2.2. Data analysis

2.2.1. Descriptive analysis

Descriptive analysis was used to present the temporal distribution of the mean temperature, relative humidity, and absolute humidity from week 14 of 2011 to week 13 of 2019 in Beijing, Tianjin, Shanxi, Hubei, Chongqing, Guangdong, Hainan, and the Hong Kong SAR, respectively (Figs. S1–S8). Subsequently, the interannual and seasonal trend of the observed weekly ILI proportion was presented with the fitted and predicted values of the GAM, GRU-based deep learning hybrid, and ARMA–GARCH models from week 14 of 2011 to week 13 of 2019 in each setting (Figs. 2–4).

2.2.2. Modelling

2.2.2.1. Methods for predicting the ILI trend using the sentinel surveillance data. The first approach is the GAM, and the basic model (Eq. (1)) is as follows:

$$log[E(Y_i)] = \alpha + ns(W_i, df) + ns(AH_i, df) + factor(H_i) + pd \quad (1)$$

where $E(Y_i)$ is the expected weekly ILI proportion in a given week *(i)*, and the link function is log(.) with the assumption that weekly ILI proportion follows normal logarithmic distribution; α is the intercept; $ns(.)$ is a cubic spline function, df is the degree of freedom; W_i is a time series comprising the number of weeks (1–52) in a calendar year, representing the potential seasonality trend in weekly ILI proportion; AH_i is the absolute humidity in the week *(i)*; and H_i is an indicator variable that equals 0–7, which represents the number of days of school or public holidays in a week *(i)*, considered as the "holiday effect" in disease reporting. The annual population density *(pd)* of each province/city was also incorporated into the model to adjust for potential demographic confounding factors across space and time. Given that the incubation period of influenza (1–4) days and the lagged effect of meteorological factors on influenza, a 1-week lag was adopted in the GAM Finally, the collinearity of the meteorological factors was analyzed before building the model, and absolute humidity was incorporated into the model (Tables S1–S8).

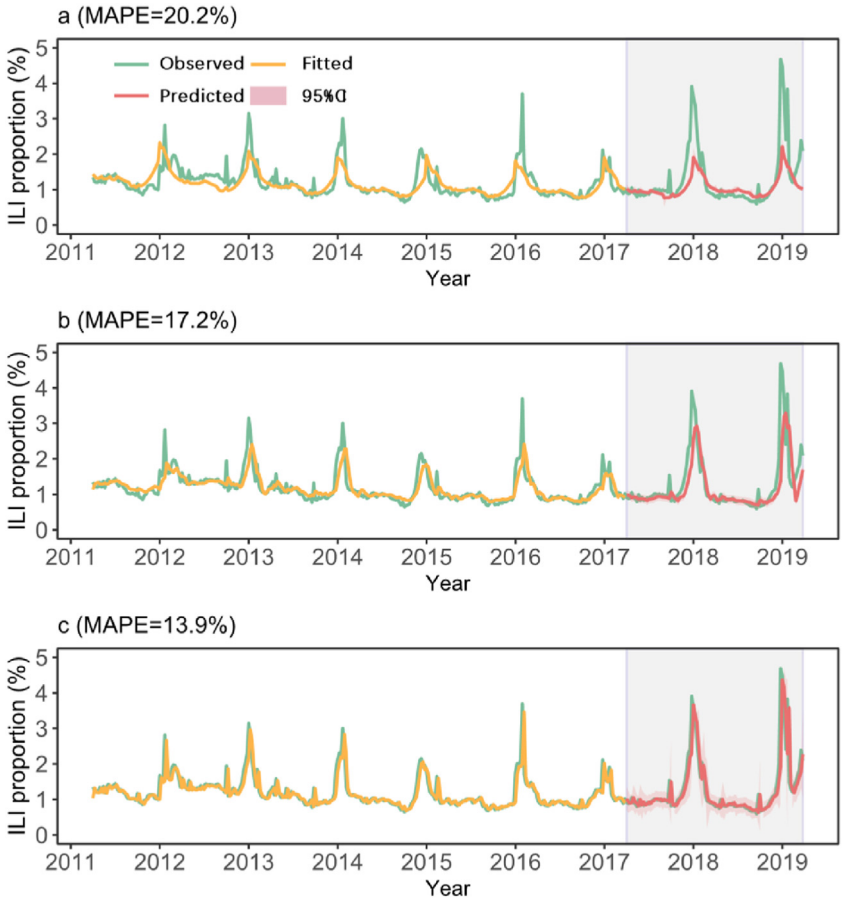

Fig. 2. Observed, fitted, and predicted weekly ILI proportion in Beijing from 2011 to 2019: (a) by the GAM; (b) GRU-based deep learning hybrid model; and (c) ARMA–GARCH model. The purple-shaded parts indicate the predicted period from week 14 in 2017 to week 13 in 2019.
Abbreviations: GAM, generalized additive model; ARMA–GARCH, autoregressive moving average-generalized autoregressive conditional; GRU, Gate Recurrent Unit; ILI, influenza-like illness.

The second approach is the deep learning hybrid model (Fig. S9), which predicts future time series by learning the features of historical time series based on deep learning. Therefore, to fully learn the characteristics and relationships of ILI historical data, we converted them into input windows of various step sizes. We built a data model based on supervised learning, which was combined with the hybrid GRU model to predict future ILI data. In this model, to enhance the model's ability to learn the features of asynchronous long data and capture the features and rules of local and overall data, we constructed data blocks 4, 8, 12, 16, 20, 24, and 28 weeks ahead based on attention mechanism concept, and subsequently fed them into the GRU module, respectively. The GRU modules perform well in processing time series and capturing the step spacing features in time series. Subsequently, the results of all branches are spliced after traversing the pooling layer to obtain the features of long unsynchronized data simultaneously. Finally, the predicted results of ILI trends were output through the dense layer.

The third approach is the ARMA–GARCH model, which is used in time series analysis, combining the ARMA and GARCH models to consider both autocorrelation and conditional heteroscedasticity in a time series. ARMA models describe the autocorrelation structure of a time series, modeling the extent to which past values of the series influence the current value. Hence, the ARMA component of the ARMA–GARCH models the autocorrelation of the time series mean. In contrast, the GARCH models describe the conditional heteroscedasticity of a time series by modeling the extent to which the variance of the series changes over time. Additionally, the GARCH component of ARMA–GARCH models the autocorrelation of the residuals (the deviations of the observed values from the mean predicted by the ARMA component). When combined, the ARMA–GARCH model provides a comprehensive description of the behavior of a time series by considering the mean and variance of the series and their dynamic relationships over time. This renders the ARMA–GARCH model a suitable tool to model ILI trends over time in this study. Further details on the time series method are available in the Supplementary Information.

When fitting and predicting using the above models, in the first step, data from week 14 of 2011 to week 13 of 2017 were used as a training set to establish models for predicting the ILI proportion in weeks 14–17 of 2017; therefore, 1-, 2-, 3-, and 4-

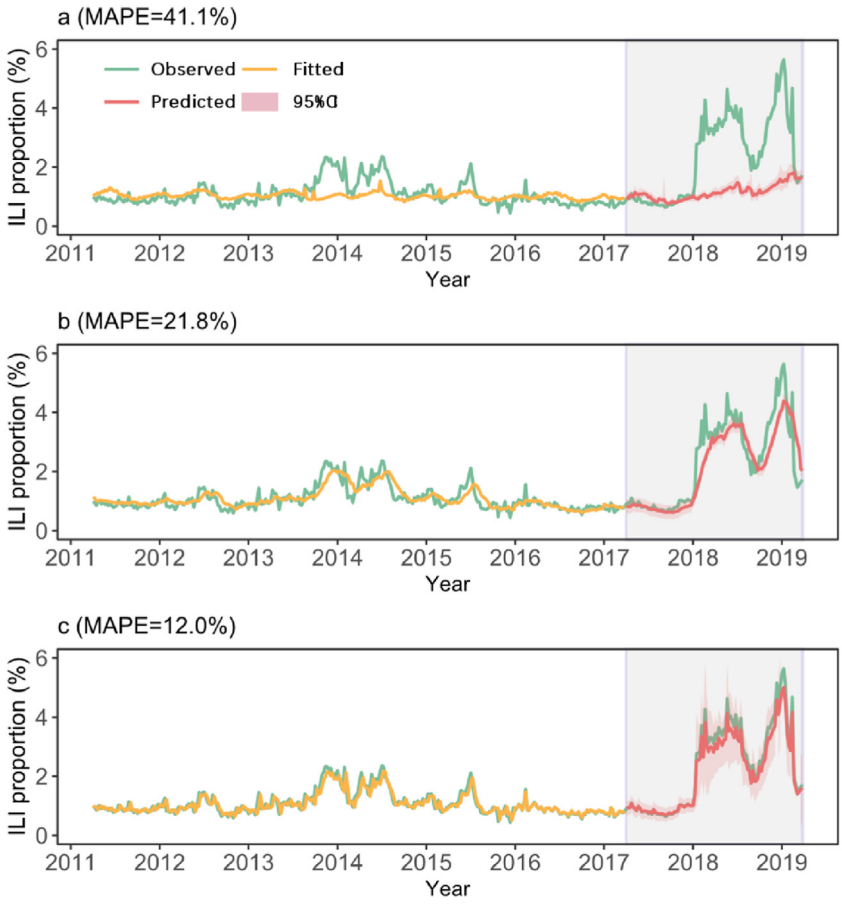

Fig. 3. Observed, fitted, and predicted weekly ILI proportion in Chongqing from 2011 to 2019: (a) by the GAM; (b) GRU-based deep learning hybrid model; and (c) ARMA–GARCH model. The purple-shaded parts indicate the predicted period from week 14 in 2017 to week 13 in 2019.
Abbreviations: GAM, generalized additive model; ARMA–GARCH, autoregressive moving average-generalized autoregressive conditional; GRU, Gate Recurrent Unit; ILI, influenza-like illness.

week-ahead predicted values of weeks 14, 15, 16, and 17 in 2017, respectively, were obtained. In the second step, data of week 14 in 2017 were added to the training set for establishing predictive models; therefore, 1-, 2-, 3-, and 4-week-ahead predicted values of weeks 15, 16, 17, and 18 in 2017, respectively, were obtained. This one-step-ahead approach continued until the ILI proportion in week 13 of 2019 was predicted. The predicted values constituted the 1-, 2-, 3-, and 4-week-ahead test sets. Additionally, the goodness of fit of the three predictive modes was assessed using Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and R squared.

2.2.2.2. Metrics for evaluating and comparing the methods. The predictive performance of the three models was measured using the three relevant indicators, which are RMSE (Eq. (2)), MAPE (Eq. (3)), and R-squared (Eq. (4)), as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[E(Y_i) - A(Y_i)]^2} \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{E(Y_i) - A(Y_i)}{A(Y_i)}\right| \tag{3}$$

$$R\ squared = 1 - \frac{\sum[A(Y_i) - E(Y_i)]^2}{\sum[A(Y_i) - \overline{A(Y_i)}]^2} \tag{4}$$

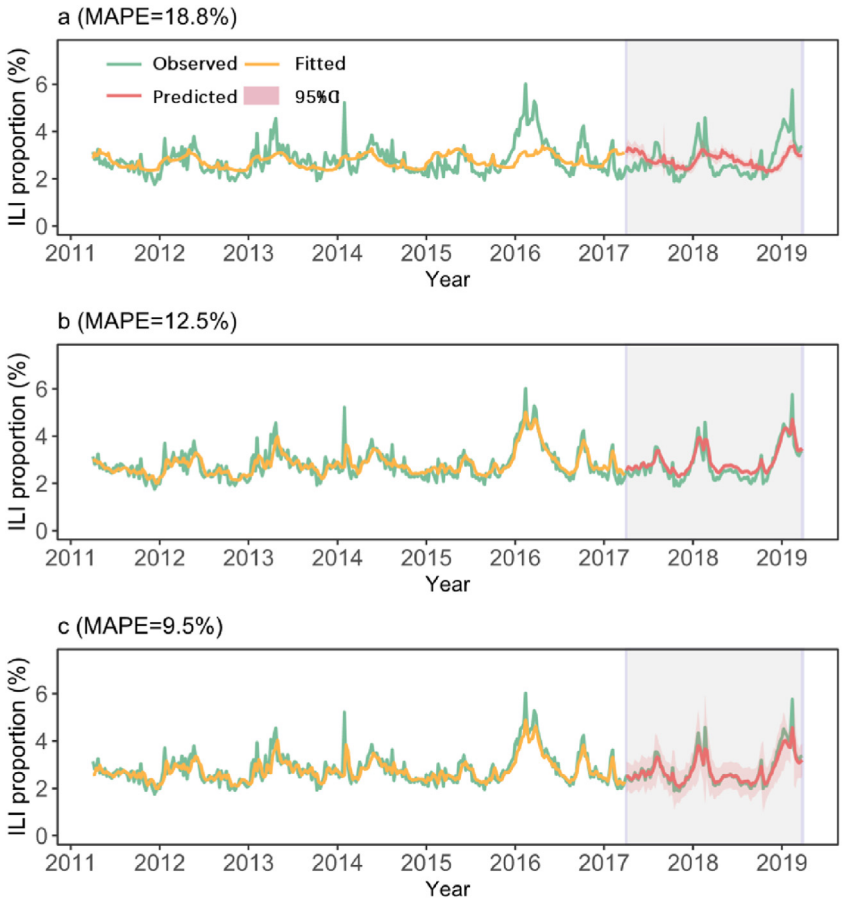

Fig. 4. Observed, fitted, and predicted weekly ILI proportion in Hainan from 2011 to 2019: (a) by the GAM; (b) GRU-based deep learning hybrid model; and (c) ARMA–GARCH model. The purple-shaded parts indicate the predicted period from week 14 in 2017 to week 13 in 2019.
Abbreviations: GAM, generalized additive model; ARMA–GARCH, autoregressive moving average-generalized autoregressive conditional; GRU, Gate Recurrent Unit; ILI, influenza-like illness.

where $E(Y_i)$ is the expected or predicted weekly ILI proportion in a given week *(i)*, which is the same as in Eq. (1); $A(Y_i)$ is the observed weekly ILI proportion in a given week *(i)*, and $\overline{A(Y_i)}$ is the mean weekly ILI proportions observed in the given period; and n represents the number of predicted values in the test set.

MAPE also represents the relative magnitude of the predicted values' deviation from the observed values and ranges from 0 to ∞. The smaller the MAPE, the better the prediction model's performance. RMSE reflects the absolute magnitude of the predicted values' deviation from the observed values, and its range is between 0 and ∞. Additionally, the smaller the RMSE, the better the prediction model's performance. According to Eqs. (2) and (3), RMSE is prone to magnifying errors, while MAPE is less susceptible to extreme values. Furthermore, R squared is an indicator evaluating the extent to which the regression model explains observed values. A low R squared value generally indicates poor performance for predictive models. However, in some cases, a good model may also show a low value. Therefore, the MAPE, RMSE, and R squared were regarded as the first, second, and assistant key metrics, respectively, in this study.

2.2.2.3. Cross-validation. A blocked cross-validation analysis (Yates et al., 2023) was performed to understand the robustness of the three predicting models in the above settings. In the first step, data from week 14 of 2011 to week 13 of 2017 (312 consecutive weeks) were used as a training set to establish the model, while those in the following 52 consecutive weeks (from week 14 of 2017 to week 13 of 2018) were used as a test set. In the second step, the dataset was shifted forward by one week for the training and test set. This analysis continued until week 13 of 2019. Finally, 53 analyses were performed for each model in each setting, and goodness of fit indicators [(mean and 95% confidence interval (CI)] were calculated.

Dataset establishment and data analyses for this study were performed using R version 4.0.0 (2020-04-24) (R Foundation for Statistical Computing, Vienna, Austria), Python version 3.6.9 (2019-7-2) (Python Software Foundation, Texas, USA), and Microsoft Excel 365MSO (207 Build 16.0.15427.20182) (Microsoft, Washington, USA).

3. Results

3.1. One-week-ahead predictive performance of the GAM, GRU-based deep learning hybrid, and ARMA–GARCH models

In temperate regions, considering Beijing as an example, the MAPE values (RMSE and R squared values are consistent with MAPE values) of the GAM, GRU-based deep learning hybrid model, and ARMA–GARCH model were 0.202, 0.172, and 0.139, respectively (Fig. 2). Overall, the ARMA–GARCH model had the best goodness of fit. In the subtropical region, with Chongqing as an example, the MAPE values of the GAM, GRU-based deep learning hybrid model, and ARMA–GARCH model were 0.411, 0.218, and 0.120 (Fig. 3), respectively, and the ARMA–GARCH model performed best. In tropical Hainan Province, the MAPE values of the GAM, GRU-based deep learning hybrid model, and ARMA–GARCH model were 0.188, 0.125, and 0.095 (Fig. 4), respectively, and the ARMA–GARCH model had the best prediction performance. Therefore, considering the goodness of fit index (Table 1), it can be inferred that the ARMA–GARCH model had the best prediction effect on ILI trend among the eight provinces/mega-cities across the three climate zones, followed by the GRU-based deep learning hybrid model with medium performance and the GAM, which had the lowest prediction accuracy.

The GAM performed best in Beijing and Shanxi, where the annual ILI peak regularly occurs in the cold winter with minimal inter-annual variation (Fig. 2a andFig. S11a). However, the GAM predicted with moderate accuracy in Tianjin (Fig. S10a), which is also located in the temperate zone but with larger inter-annual variations. Furthermore, the GAM performed poorly in Hubei (Fig. S12a), Chongqing (Fig. 3), Guangdong Fig. S13a), Hainan (Fig. 4), and the Hong Kong SAR Fig. S14a).

The predicted peaks of the GRU-based deep learning hybrid model lagged behind those observed by 1–3 weeks in almost every influenza season in most settings (Figs. 2–4, Figs. S10–S14), besides its moderate predictive performance.

3.2. Two-, three-, and four-week-ahead predictive performance of the GAM, GRU-based deep learning hybrid, and ARMA–GARCH model

The prediction performance of GAM gradually decreased as the number of weeks ahead increased in the eight provinces/megacities Tables S9–S11). In temperate zones, with Beijing as an example, the MAPE values predicted 1-, 2-, 3-, and 4-week-ahead were 0.202, 0.206, 0.209, and 0.212, respectively. However, the values predicted 1-, 2-, 3-, and 4-week-ahead in the subtropical regions, considering Chongqing as an example, were 0.411, 0.423, 0.434, and 0.442, respectively. In tropical Hainan Province, the MAPE values predicted 1-, 2-, 3-, and 4-week-ahead were 0.188, 0.193, 0.195, and 0.197, respectively.

The GRU-based deep learning hybrid model's prediction performance also gradually declined as the number of weeks ahead increased, with a few exceptions in some provinces (Table 1). In temperate zones, with Beijing as an example, the MAPE values predicted 1-, 2-, 3-, and 4-week-ahead were 0.172, 0.181, 0.191, and 0.203, respectively. However, the MAPE values predicted 1-, 2-, 3-, and 4-week-ahead were 0.218, 0.260, 0.293, and 0.332, respectively, in subtropical regions, with Chongqing as an example. Furthermore, in tropical Hainan Province, the MAPE values predicted 1-, 2-, 3-, and 4-week-ahead were 0.125, 0.157, 0.176, and 0.178, respectively. Remarkably, in the Hubei Province, the prediction performance of 4-week-ahead was better than that of 3-weeks-ahead (MAPE = 0.272 and 0.256, respectively).

Similar to that of the GAM, the prediction accuracy of the ARMA–GARCH model in the eight provinces/megacities decreased with an increased number of weeks ahead. In temperate zones, considering Beijing as an example, the MAPE values predicted 1-, 2-, 3-, and 4-week-ahead were 0.139, 0.184, 0.234, and 0.269, respectively. The values predicted 1-, 2-, 3-, and 4-week-ahead in the subtropical region of Chongqing were 0.120, 0.177, 0.225, and 0.261, respectively. Moreover, in tropical Hainan Province, the MAPE values predicted 1-, 2-, 3-, and 4-week-ahead were 0.095, 0.120, 0.138, and 0.149, respectively.

Therefore, a comparison of the three models' prediction performance revealed that the results of 2-, 3-, and 4-week-ahead results were similar to that of 1-week-ahead for most provinces/megacities (Table 1, Tables S9–S11). The ARMA–GARCH model had the best prediction performance, while GAM performed poorly, and the GRU-based deep learning hybrid model had a moderate performance. However, there were a few exceptions, such as the 2-weeks-ahead prediction performance of the GAM in Tianjin, which was better than that of the GRU-based deep learning hybrid model (MAPE values were

Table 1
Predictive performance of the three models 1-week-ahead in eight provinces/megacities, 2011–2019.

	GAM			Deep learning hybrid model			ARMA–GARCH		
	MAPE	RMSE	R squared	MAPE	RMSE	R squared	MAPE	RMSE	R squared
Beijing	0.202	0.668	0.377	0.172	0.566	0.552	0.139	0.470	0.698
Tianjin	0.256	1.783	−0.089	0.205	1.094	0.591	0.102	0.734	0.819
Shanxi	0.162	0.391	0.610	0.157	0.445	0.494	0.127	0.327	0.736
Hubei	0.381	1.034	−0.745	0.155	0.546	0.511	0.087	0.356	0.792
Chongqing	0.411	1.746	−0.520	0.218	0.767	0.705	0.120	0.533	0.882
Guangdong	0.160	1.092	−0.389	0.134	0.749	0.347	0.082	0.593	0.616
Hainan	0.188	0.668	0.114	0.125	0.443	0.612	0.095	0.419	0.666
Hong Kong SAR	0.228	1.212	0.100	0.223	1.029	0.346	0.182	0.887	0.515

GAM, generalized additive model; ARMA–GARCH, autoregressive moving average-generalized autoregressive conditional; GRU, Gate Recurrent Unit; MAPE, Mean Absolute Percentage Error; RMSE, Root Mean Squared Error; SAR, Special Administrative Region.

0.261 and 0.330, respectively). The 3-weeks-ahead prediction performance of GAM in the Hong Kong SAR was superior to that of the GRU-based deep learning hybrid model (MAPE values were 0.297 and 0.319, respectively).

3.3. Robustness of the GAM, GRU-based deep learning hybrid, and ARMA−−GARCH models

The three models' prediction performances in each block were relatively stable (reasonable mean values of RMSE, MAPE, and R squared with lower 95% CIs), as per the blocked cross-validation analysis (Table 2). This further implies that all models were robust to the changes in data and had a low risk of overfitting.

4. Discussion

Influenza is a contagious respiratory illness caused by influenza viruses that has a significant disease burden globally. The transmission pattern of influenza varies across climate zones globally (Azziz Baumgartner et al., 2012). ILI is a sensitive proxy indicator of influenza activity that is key to understanding the influenza trend. Additionally, due to the similarity in symptoms, such as cough and fever, ILI can provide insights into the trends of other respiratory diseases (Aung et al., 2021). Therefore, this study evaluated and compared the prediction performance of three models, which are the GAM, GRU-based deep learning hybrid model, and ARMA−GARCH model, in predicting ILI trend and found that the ARMA−GARCH model attained the best accuracy, followed by the GRU-based deep learning hybrid model irrespective of the climate zone.

The GRU-based deep learning hybrid model addresses the challenges of gradient explosion or disappearance commonly encountered in simpler recurrent neural networks (He et al., 2022; Tang Chaofan et al., 2023). This model is particularly adept at predicting significant events in time series data characterized by long intervals and substantial time delays. A notable application is in the ILI sentinel surveillance data, where the model demonstrates considerable predictive capability. One of the distinct features of the model is its multi-step input mode, which allows it to capture both macro and micro temporal features within the ILI time series. This capability enables a more nuanced understanding and modeling of the data dynamics, which is critical for accurate forecasting. However, despite its strengths, the model exhibits some limitations in the timing of its predictions. Specifically, the predicted peaks for ILI occurrences tend to lag behind the observed ones by approximately 1−3 weeks. This lag indicates potential areas for improvement in the model's architecture or training process to enhance its responsiveness and accuracy in real-time forecasting scenarios. To address the observed lag in peak prediction, future work could explore the integration of additional predictive indicators or the implementation of real-time data adjustment techniques. Also, experimenting with different configurations of the GRU layers or introducing complementary neural network structures might refine the model's predictive performance. Nevertheless, the model excels in handling extremely large-scale datasets (Reichstein et al., 2019), making it particularly suitable for applications where vast amounts of data need to be processed efficiently. Notably, the conclusions of this study are primarily applicable under the current conditions and with the specific data characteristics involved in this research.

In the realm of epidemiological analysis, the ARMA-GARCH model offers a refined methodology for modeling and forecasting time series data that exhibits periodic trends and volatility clustering, such as those seen in ILI trends in the study. The ARMA (Autoregressive Moving Average) component of the model adeptly captures the linear dependencies and periodic fluctuations in the data, employing autoregressive and moving average terms to accurately describe the conditional mean of the series (Box et al., 2015). Complementing this, the GARCH (Generalized Autoregressive Conditional Heteroskedasticity)

Table 2
Goodness of fit of the three models with blocked cross-validation method in eight provinces/megacities, 2011−2019.

	GAM			deep learning hybrid model			ARMA−GARCH		
	RMSE	MAPE	R squared	RMSE	MAPE	R squared	RMSE	MAPE	R squared
Beijing	0.416 (0.399, 0.433)	49.186 (37.709, 60.663)	0.131 (0.042, 0.221)	0.476 (0.451, 0.501)	0.149 (0.145, 0.153)	0.642 (0.623, 0.662)	0.374 (0.347, 0.402)	0.127 (0.124, 0.130)	0.780 (0.760, 0.801)
Tianjin	0.413 (0.384, 0.442)	0.239 (0.221, 0.258)	−0.474 (−0.835, −0.112)	0.230 (0.215, 0.245)	1.025 (0.986, 1.064)	0.230 (0.215, 0.245)	0.095 (0.091, 0.099)	0.586 (0.542, 0.629)	0.095 (0.091, 0.099)
Shanxi	0.203 (0.198, 0.209)	2.406 (2.292, 2.521)	0.679 (0.649, 0.710)	0.441 (0.429, 0.453)	0.163 (0.165, 0.161)	0.542 (0.523, 0.561)	0.308 (0.301, 0.315)	0.132 (0.129, 0.136)	0.778 (0.760, 0.797)
Hubei	0.371 (0.341, 0.401)	0.439 (0.383, 0.495)	−1.404 (−1.641, −1.167)	0.165 (0.160, 0.169)	0.501 (0.485, 0.516)	0.165 (0.160, 0.169)	0.097 (0.095, 0.099)	0.338 (0.326, 0.350)	0.097 (0.095, 0.099)
Chongqing	1.071 (1.007, 1.136)	–	−7.737 (−10.337, −5.137)	0.200 (0.193, 0.207)	0.732 (0.704, 0.761)	0.200 (0.193, 0.207)	0.115 (0.111, 0.119)	0.516 (0.496, 0.536)	0.115 (0.111, 0.119)
Guangdong	0.374 (0.343, 0.405)	0.442 (0.385, 0.499)	−1.448 (−1.696, −1.201)	0.131 (0.129, 0.132)	0.686 (0.672, 0.700)	0.131 (0.129, 0.132)	0.074 (0.072, 0.076)	0.492 (0.470, 0.513)	0.074 (0.072, 0.076)
Hainan	0.222 (0.217, 0.227)	0.212 (0.208, 0.216)	−0.103 (−0.124, −0.083)	0.136 (0.134, 0.137)	0.442 (0.439, 0.446)	0.136 (0.134, 0.137)	0.092 (0.090, 0.093)	0.395 (0.388, 0.401)	0.092 (0.090, 0.093)
Hong Kong SAR	0.379 (0.361, 0.396)	0.274 (0.248, 0.299)	−0.690 (−0.794, −0.586)	0.241 (0.231, 0.252)	1.104 (1.092, 1.115)	0.241 (0.231, 0.252)	0.192 (0.183, 0.202)	0.093 (0.091, 0.094)	0.192 (0.183, 0.202)

GAM, generalized additive model; ARMA−GARCH, autoregressive moving average-generalized autoregressive conditional; GRU, Gate Recurrent Unit; MAPE, Mean Absolute Percentage Error; RMSE, Root Mean Squared Error; SAR, Special Administrative Region.

component specifically addresses the conditional variance, effectively modeling the dynamic and often unpredictable changes in volatility that can occur due to environmental factors, social interactions, and other external influences (Bollerslev, 1986). This combined modeling approach significantly enhances the capability to predict and understand disease trends, providing vital insights for public health planning and intervention strategies.

With an increase in the number of weeks ahead, the prediction performance of the GAM, GRU-based deep learning hybrid model, and ARMA−−GARCH model gradually worsened, with the performance of 1-week-aheadbest for all three models. This phenomenon is related to the models' internal operation patterns andILI's transmission patterns. That is, ILI comprises a series of symptoms, such as fever of $\geq$38 °C and cough, which may be caused by influenza, severe acute respiratory syndrome coronavirus 2, rhinovirus, parainfluenza virus, respiratory syncytial virus, and many other respiratory viral infections. Respiratory infectious diseases spread through the respiratory tract, close contact, and other means, casusing the spatial and temporal distribution of cases to be interrelated but not independent (Leite et al., 2021), implying that the previous week's data are closely related to that of the following week. Therefore, whether the previous week's data are included or excluded in the training set of the model significantly impacts the prediction accuracy of the specific week. Additionally, the more the number of weeks ahead, the lower the prediction accuracy, which implies a worse prediction performance. However, the prediction performance of the GRU-based deep learning hybrid model occasionally out-performed those of the GAM and ARMA-GARCH models with an increase in the number of weeks ahead, which may be related to the waveform of ILI trend that the model mainly learns (Tian et al., 2020). Meanwhile, its prediction stability also declines when the waveform of a certain province or city is unstable (considerable clutter).

This study has some limitations. First, excluding certain temperature settings, the prediction accuracy of the GAM was remarkably poor. The more complex circulation pattern of influenza in subtropical or tropical zones could partially account for this phenomenon. However, the main nonparametric terms incorporated into the GAM in this study were the absolute humidity and number of weeks in a calendar, which are key factors that drive the seasonal variations of influenza. Therefore, to attain better prediction performance in the future, other drivers should be explored and considered in the GAM, particularly in subtropical and tropical settings, such as population mobility or density. Second, the lag times of 1-, 2-, 3-, and 4-week-ahead were used, which could have improved the prediction accuracies compared to longer lag times. However, forecasting ILI trends several weeks ahead, such as 8 or 16 weeks, can help researchers gain early insights into influenza trends to facilitate preparations for influenza epidemics and outbreak responses. Therefore, striking a balance when deciding the lag time in prediction is crucial in public health practice. Third, due to the data accessibility, the study period ranged from 2011 to 2019, excluding the COVID-19 pandemic. Therefore, the influenza transmission patterns may have been impacted, or even partly changed, by the COVID-19 pandemic and related NPIs. Therefore, future studies focused on influenza prediction during or after the COVID-19 pandemic are warranted to enrich existing evidence.

Collectively, findings of the study shed light on the influenza prediction approaches, which inform the application of the ARMA-GARCH model to predict disease epidemic trends or optimize relevant polices, particularly under the scenario of short-time forecast (e.g., 1-week ahead).

5. Conclusions

The ARMA–GARCH model exhibited the best accuracy in predicting ILI trends in China, compared with the GAM and GRU-based deep learning hybrid model, irrespective of the number of weeks ahead, whether it was 1-, 2-, 3-, or 4-week-ahead. Therefore, based on our findings, the ARMA–GARCH model may be considered for predicting ILI trends in public health practice in varying climate zones. Its implementation can contribute to enhanced efforts in influenza control and prevention in the future.

Funding

The Special Fund for Health Development Research of Beijing (2021-1G-3013), the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (2021-I2M-1-044), and the Bill & Melinda Gates Foundation (INV-024911).

Ethics approval and consent to participate

The ILI data analyzed in the study were collected as part of routine surveillance program and no individual information was involved, so it could be exempt from ethics approval.

Consent for publication

Not applicable.

Data statement

The ILI datasets in Beijing, Tianjin, Shanxi, Hubei, Chongqing, Guangdong and Hainan analyzed during the current study are available from the corresponding authors on a reasonable request. The ILI datasets in the Hong Kong SAR analyzed during

the current study are available from the Centre for Health Protection of Hong Kong SAR portal, https://www.chp.gov.hk/en/static/24015.html. The meteorological datasets in the eight settings analyzed in the current study are available from China meterological data service center, http://data.cma.cn/en, and the National Centers for Environmental Information portal, https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-day?bbox=53.544,73.620,18.198,134.761&place = Country:194&stations = 54511099999&pageNum = 4. The public and school holiday datasets analyzed during the current study are available from the WorldPop Data portal, https://hub.worldpop.org/project/categories?id=19. The yearly population density datasets of the eight settings analyzed during the current study are available from the statistical yearbooks from Beijing Municipal Bureau Statistics portal, https://nj.tjj.beijing.gov.cn/nj/main/2022-tjnj/zk/indexch.htm, Tianjin Municipal Bureau Statistics portal, https://stats.tj.gov.cn/, Shanxi Provincial Statistics portal, http://tjj.shanxi.gov.cn/, Hubei Provincial Bureau Statistics portal, http://tjj.hubei.gov.cn/, Chongqing Municipal Bureau Statistics portal, http://tjj.cq.gov.cn/, Guangdong Provincial Bureau Statistics portal, http://stats.gd.gov.cn/, Hainan Provincial Bureau Statistics portal, http://stats.hainan.gov.cn/tjj/index.html, and Census and Statistics Department of the Government of the Hong Kong SAR portal, https://www.censtatd.gov.hk/sc/EIndexbySubject.html?pcode=B1120017&scode=150.

CRediT authorship contribution statement

Xingxing Zhang: Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Liuyang Yang:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Teng Chen:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Qing Wang:** Data curation. **Jin Yang:** Visualization, Software, Resources. **Ting Zhang:** Visualization, Software, Resources. **Jiao Yang:** Visualization, Software, Resources. **Hongqing Zhao:** Visualization, Software, Resources. **Shengjie Lai:** Methodology, Conceptualization. **Luzhao Feng:** Writing – review & editing, Methodology, Conceptualization. **Weizhong Yang:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

We thank all the clinicians and patients who provided valuable ILI data for the study. We thank Editage (app.editage.cn) for the English language editing and review services.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2024.04.010.

References

Adlhoch, C., Sneiderman, M., Martinuka, O., Melidou, A., Bundle, N., Fielding, J., Olsen, S. J., Penttinen, P., Pastore, L., & Pebody, R. (2021). Spotlight influenza: The 2019/20 influenza season and the impact of COVID-19 on influenza surveillance in the WHO European Region. *Euro Surveillance, 26*(40). https://doi.org/10.2807/1560-7917.ES.2021.26.40.2100077

Aung, A. H., Lye, D. C., Cui, L., Ooi, C. K., & Chow, A. L. P. (2021). The 'timeless' use of influenza-like illness criteria for influenza detection in the tropics. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases, 106*, 160–168. https://doi.org/10.1016/j.ijid.2021.03.045

Azziz Baumgartner, E., Dao, C. N., Nasreen, S., Bhuiyan, M. U., Mah-E-Muneer, S., Mamun, A. A., Sharker, M. Y., Zaman, R. U., Cheng, P. Y., Klimov, A. I., & Widdowson, M. A. (2012). Seasonality, timing, and climate drivers of influenza activity worldwide. *Journal of Infectious Diseases, 206*(6), 838–846. https://doi.org/10.1093/infdis/jis467

Beijing Municipal Bureau Statistics. Beijing Statistical Yearbook. Retrieved from http://tjj.beijing.gov.cn/. Accessed November 1, 2022.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

Census and Statistics Department of the Government of the Hong Kong Special Administration Region. Population Census. Retrieved from https://www.censtatd.gov.hk/sc/. Accessed November 1, 2022.

Center for Health Protection/Department of Health of the Government of the Hong Kong Special Administrative Region. (2022b). General out-patient clinics/private medical practitioner clinics. https://www.chp.gov.hk/en/static/24015.html.

Centers for Disease Control and Prevention. (2023). *Influenza surveillance report*. Retrieved from https://www.cdc.gov/flu/weekly/index.htm. (Accessed 1 March 2023).

China Meteorological Administration. China meteorological data service center. http://data.cma.cn/en. (Accessed 1 November 2022).

Chongqing Provincial Bureau Statistics. Chongqing statistical yearbook. http://tjj.cq.gov.cn/. (Accessed 1 November 2022).

Cowling, B. J., Ali, S. T., Ng, T. W. Y., Tsang, T. K., Li, J. C. M., Fong, M. W., Liao, Q., Kwan, M. Y., Lee, S. L., Chiu, S. S., & Wu, J. T. (2020). Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: An observational study. *The Lancet Public Health, 5*(5), e279–e288. https://doi.org/10.1016/S2468-2667(20)30090-6

Cowling, B. J., Ho, L. M., Riley, S., & Leung, G. M. (2013). Statistical algorithms for early detection of the annual influenza peak season in Hong Kong using sentinel surveillance data. *Hong Kong Medical Journal, 19*(Suppl. 4), 4–5.

Feng, L., Zhang, T., Wang, Q., Xie, Y., Peng, Z., Zheng, J., Qin, Y., Zhang, M., Lai, S., Wang, D., & Feng, Z. (2021). Impact of COVID-19 outbreaks and interventions on influenza in China and the United States. *Nature Communications, 12*(1), 3249. https://doi.org/10.1038/s41467-021-23440-1

Guangdong Provincial Bureau Statistics. Guangdong statistical yearbook. http://stats.gd.gov.cn/. (Accessed 1 November 2022).

Hainan Municipal Bureau Statistics. Hainan Statistical Yearbook. Retrieved from http://stats.hainan.gov.cn/tjj. Accessed November 1, 2022[.

Hay, A. J., & McCauley, J. W. (2018). The WHO global influenza surveillance and response system (GISRS)-A future perspective. *Influenza and Other Respiratory Viruses, 12*(5), 551–557. https://doi.org/10.1111/irv.12565

He, Y., Lu, Z., Wang, W., Zhang, D., Zhang, Y., Qin, B., Shi, K., & Yang, X. (2022). Water clarity mapping of global lakes using a novel hybrid deep-learning-based recurrent model with Landsat OLI images. *Water Research, 215*, Article 118241. https://doi.org/10.1016/j.watres.2022.118241

Lai, S., Sorichetta, A., Steele, J., Ruktanonchai, C. W., Cunningham, A. D., Rogers, G., Koper, P., Woods, D., Bondarenko, M., Ruktanonchai, N. W., & Shi, W. (2022). Global holiday datasets for understanding seasonal human mobility and population dynamics. *Scientific Data, 9*(1), 17. https://doi.org/10.1038/s41597-022-01120-z

Lau, E. H. Y., Cowling, B. J., Ho, L. M., & Leung, G. M. (2008). Optimizing use of multistream influenza sentinel surveillance data. *Emerging Infectious Diseases, 14*(7), 1154–1157. https://doi.org/10.3201/eid1407.080060

Lazarus, J. V., Romero, D., Kopka, C. J., Karim, S. A., Abu-Raddad, L. J., Almeida, G., Baptista-Leite, R., Barocas, J. A., Barreto, M. L., Bar-Yam, Y., & Bassat, Q. (2022). A multinational Delphi consensus to end the COVID-19 public health threat. *Nature, 611*(7935), 332–345. https://doi.org/10.1038/s41586-022-05398-2

Leite, G. S., Albuquerque, A. B., & Pinheiro, P. R. (2021). Applications of technological solutions in primary ways of preventing transmission of respiratory infectious diseases-A systematic literature review. *International Journal of Environmental Research and Public Health, 18*(20). https://doi.org/10.3390/ijerph182010765

National Health Commission of the People's Republic of China. (2017). *Notice on printing and distributing the national influenza surveillance program*. Edition. Retrieved from http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=ed1498d9e64144738cc7f8db61a39506.

National Oceanic and Atmospheric Administration. National Centers for Environmental Information. Retrieved from https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-day?bbox=53.544,73.620,18.198,134.761&place=Country:194&stations=54511099999&pageNum=4. Accessed November 1, 2022.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature, 566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Shaman, J., & Kohn, M. (2009). Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America, 106*(9), 3243–3248. https://doi.org/10.1073/pnas.0806852106

Shanxi Provincial Bureau Statistics. Shanxi statistical yearbook. http://tjj.shanxi.gov.cn/. (Accessed 1 November 2022).

Tang Chaofan, L. X., Yang, B., Tang, Y., & Zhao, D. (2023). GRU-based interpretable multivariate time series anomaly detection in industrial control system. *Computers & Security, 127*, Article 103094.

Tian, C., Fei, L., Zheng, W., Xu, Y., Zuo, W., & Lin, C. W. (2020). Deep learning on image denoising: An overview. *Neural Networks: The Official Journal of the International Neural Network Society, 131*, 251–275. https://doi.org/10.1016/j.neunet.2020.07.025

Tianjin Municipal Bureau Statistics. Tianjin Statistical Yearbook. Retrieved from https://stats.tj.gov.cn/. Accessed November 1, 2022.

Uyeki, T. M., Hui, D. S., Zambon, M., Wentworth, D. E., & Monto, A. S. (2022). Influenza. *Lancet, 400*(10353), 693–706. https://doi.org/10.1016/S0140-6736(22)00982-5

World Health Organization. Global influenza surveillance and response system (GISRS). Retrieved from https://www.who.int/initiatives/global-influenza-surveillance-and-response-system.

WorldPop. (2022). WorldPop data. https://www.worldpop.org/datacatalog/. (Accessed 1 November 2022).

Wu, P., Presanis, A. M., Bond, H. S., Lau, E. H. Y., Fang, V. J., & Cowling, B. J. (2017). A joint analysis of influenza-associated hospitalizations and mortality in Hong Kong, 1998–2013. *Scientific Reports, 7*(1), 929. https://doi.org/10.1038/s41598-017-01021-x

Yang, W., Cowling, B. J., Lau, E. H., & Shaman, J. (2015). Forecasting influenza epidemics in Hong Kong. *PLoS Computational Biology, 11*(7), Article e1004383. https://doi.org/10.1371/journal.pcbi.1004383

Yates, L. A., Aandahl, Z., Richards, S. A., & Brook, B. W. (2023). Cross validation for model selection: A review with examples from ecology. *Ecological Monographs, 93*(1), Article e1557. https://doi.org/10.1002/ecm.1557

Yu, H., Alonso, W. J., Feng, L., Tan, Y., Shu, Y., Yang, W., & Viboud, C. (2013). Characterization of regional influenza seasonality patterns in China and implications for vaccination strategies: Spatio-temporal modeling of surveillance data. *PLoS Medicine, 10*(11), Article e1001552. https://doi.org/10.1371/journal.pmed.1001552