Dean&Francis
ISSN 2959-409X

Seasonal Influenza Transmission Patterns and Predictive Modeling

Yunkai Zhang

Department of Biological Sciences,
University of California, Irvine,
California, USA
Corresponding author: yunkaiz2@uci.edu

Abstract:

Seasonal influenza is a recurring public-health burden that peaks in winter and tests outpatient and hospital capacity. Accurately predicting its seasonality and near-term activity is crucial to preventing avoidable hospitalisations and activating and timing season-appropriate preparedness actions. This article distils salient evidence on environmental drivers and data streams for multi-faceted surveillance and operational prediction, as well as modelling families that complement the latter. It overviews laboratory and epidemiologic evidence that absolute humidity modulates viral survival and transmissibility, thus partly explaining influenza's robust winter seasonality in temperate climates. It then discusses surveillance systems ranging from the global WHO FluNet/GISRS to national dashboards such as CDC FluView and ECDC ERVISS that provide timely, standardised indicators and prediction targets. It subsequently reviews main method families (statistical time-series, mechanistic compartmental models, machine learning) and highlights that multi-model ensembles often outperform single approaches at predicting onset, peak week, and intensity. It also presents case studies from Europe, the U.S. and China that illustrate prediction skill, the importance of climate-aware forcing, and of practical factors such as reporting lags. Generally, robust and interpretable forecasts require well-aligned targets, hybrid/ensemble frameworks, adaptive weighting and drift monitoring, and hold promise to support earlier warnings, intervention scenario analysis, and efficient resource planning.

Keywords: Seasonal influenza; absolute humidity; surveillance systems; SEIR modeling; ensemble forecasting.

1. Introduction

Seasonal influenza represents a recurrent burden to population health as annual epidemics lead to substantial morbidity and mortality and cause peaks in outpatient attendances, emergency department overcrowding and hospital admissions, which repeatedly test healthcare services. Furthermore, the burden is distributed unevenly by age and clinical risk group, and for example, U.S. CDC's FluSurv-NET—which is part of U.S. influenza surveillance and monitors laboratory-confirmed influenza-associated hospitalisations—reveals that young children (<5 years) and older adults (>65 years) are overrepresented in severe cases while working-age adults drive transmission and overall case incidence, which complicates preparation and resource allocation efforts [1,2]. A well-defined seasonal pattern, when paired with reliable prediction, would help to curb avoidable hospitalisations and facilitate early implementation of preparatory actions, and this justifies the need for understanding influenza transmission mechanisms and developing predictive models. Research in this regard depends on good surveillance data, and on the global level, the WHO Global Influenza Surveillance and Response System (GISRS)—which includes national influenza centres, influenza reference laboratories, and collaborating laboratories in more than 130 WHO member states—serves as the basis of surveillance with FluNet, its online platform collating weekly virologic data (including test counts and subtype-positive results) for real-time situational awareness and later analysis [1]. At the national and subcontinental levels, systems like CDC FluView in the U.S.—which is a composite of laboratory, outpatient ILI, and hospital data—and ERVISS, the European Centre for Disease Prevention and Control's comprehensive dashboard for influenza, RSV, and SARS-CoV-2, further contribute to a hierarchical pipeline that provides data streams (e.g., weekly incidence, age-stratified rates) to define key features of the epidemic and model components such as epidemic onset and peak week [1-3].

Seasonal influenza transmission is highly seasonal in most temperate regions, with environmental conditions systematically and predictably modulating the underlying biological and behavioural processes. Absolute humidity in particular is a key factor for virus survival and transmissibility, as low absolute humidity during winter promotes optimal environmental persistence of virions and efficient respiratory transmission that drives the winter peaks observed in most countries; examples include U.S. activity peaking December–February and European epidemics having winter dominance with wider "shoulder" phases, and this informs climate-informed modeling [2-4]. Exceptions exist in transitional or subtropical regions due to climate heterogeneity or social behavior differences, and FluNet data also reveal interannual variability, including skipped seasons and subtype-specific variations [1]. Forecasting capability for seasonal influenza has matured along three mutually reinforcing families of methods, with statistical time-series models such as autoregressive integrated moving average with seasonal variations enabling rapid and reasonably precise short-horizon forecasts when surveillance data streams are stable. Mechanistic compartmental models such as the susceptible–exposed–infectious–recovered (SEIR) model explicitly encode the transmission process to allow counterfactual exploration of intervention impacts, including vaccination and antiviral use, while ensemble systems combine two or more models—often integrating statistical, mechanistic, and machine-learning approaches—to mitigate bias and variability of individual models. Multi-season independent evaluations have confirmed that multi-model ensembles enhance precision for standard public-health targets, including epidemic onset, peak week, and peak intensity, which supports their routine use in surveillance programmes [5]. This paper synthesizes the latest evidence on seasonal influenza transmission and predictive modeling, integrating surveillance-consistent targets and data structures, comparing time-series, SEIR, and machine-learning techniques in terms of typical use cases, performance characteristics, and interpretability, and offering practical advice for model evaluation procedures, application of climate signals, including absolute humidity, and multi-model ensemble design. Ultimately, it aims to provide a brief and transferable toolkit that facilitates short-term prediction and scenario planning across diversified data settings

2. Analysis of Seasonal Influenza Transmission Patterns

2.1 Transmission Drivers and Seasonality

A well-documented empirical observation is that influenza exhibits strong seasonality, particularly in temperate zones worldwide, and a long-standing question in modeling has been how to mechanistically incorporate seasonality rather than merely fitting it as a function to data. In this context, absolute humidity (AH) emerges as a leading environmental variable, and reanalysis of multiple laboratory and epidemiological datasets has found that AH explained ~90% and ~50% of the variability in influenza virus survival (IVS) and transmission efficiency (IVT) respectively while relative humidity (RH) explained only ~36% and ~12% [4]. The proposed mechanism is that lower AH in winter reduces humidity-driven stress on viral particles,

ISSN 2959-409X

prolonging their airborne persistence and thereby increasing transmission. Beyond laboratory-scale correlations, extension of this finding to human epidemic dynamics has demonstrated that anomalously low AH in the weeks preceding an influenza outbreak is associated with the timing of winter influenza epidemics in the U.S. A model experiment in which AH was used to modulate the time-varying basic reproductive number ($R_0(t)$) successfully recapitulated the timing and magnitude of observed mortality peaks, which suggests that AH forcing itself can account for a large portion of influenza seasonality in temperate zones and supports the hypothesis that environmental modulation of transmissibility is not merely a correlation but a proximal driver of influenza seasonality [4].

However, several important caveats require emphasis. The AH-based modulation in the aforementioned studies assumes that population-internal factors, including immune waning, contact dynamics, and virus importations, are secondary or subordinate, but this simplification may not hold in practice, as these factors can interact nonlinearly with climatic forcing. The importance of AH is also context-specific, depending on location, local climate regime, and indoor/outdoor settings; an example is that AH measures water content in air, and indoor humidity in heated spaces can become decoupled from ambient AH. Additionally, while AH is a strong predictor in temperate regions, it is less explanatory in tropical or highly humid regions where AH varies minimally. Despite these qualifications, the evidence from the aforementioned reanalysis is widely regarded as foundational, leading to the inclusion of AH or humidity-based forcing functions in forecasting systems primarily for temperate environments [6].

2.2 Clinical Burden and Surveillance Limitations

Surveillance systems like FluSurv-NET and FluView offer rich resolution on the influenza burden (e.g., weekly age-stratified, region-specific incidence of hospitalizations). However, the time taken for laboratory confirmation (needed for objective diagnosis of influenza) and non-uniform geographic surveillance coverage still result in lag, under-ascertainment, and data incompleteness. These data limitations therefore circumscribe the development, evaluation, and testing of forecasting systems; for instance, a model learned in one area from complete real-time data may not be as effective in a data-limited area where lags are present. Additionally, the age-stratified burden—with severe cases concentrated in <5 and >65 years—informs not only transmission dynamics, including working-age adults as key transmitters, but also model design, such as including age-specific compartments in mechanistic models to improve predictive accuracy [2].

3. Analysis of Predictive Modeling Methods

3.1 Mechanistic Models and Extension

Mechanistic compartmental models such as SEIR and SEIQD offer both interpretability and flexibility to integrate environmental forcing and intervention measures, supporting counterfactual simulations, including assessing vaccination or antiviral uptake and incorporating time-varying transmissibility. A recent example of such a model is a SEIR-type model "designed to infer influenza vaccine efficacy in a real-world setting using surveillance data from the Valencian Community, Spain"[7]. This model segments the population into vaccinated and unvaccinated compartments, linking them with differential transmission dynamics, and includes a vaccine attenuation factor (a_v) that scales down infection risk in the vaccinated population, while the fitting process jointly estimates the transmission rate (β), vaccine attenuation (a_v) and an auxiliary parameter (k)—which does not alter internal dynamics but is used to fit total error. A main conclusion of this work is that the mechanistic structure can be used flexibly and tuned to estimate real vaccine performance, rather than simulating candidate interventions.

Several other important conclusions may also be drawn from this study. This work shows that mechanistic models do not necessarily have to be idealized, and that they can be fit to real-world data (including those with intervention effects) within a single influenza season. The calibration issue (e.g. non-identifiability between β, a_v and k) identified here suggests that realistic mechanistic models with sensible constraints or priors on parameters are needed to prevent overfitting, and the success of this model in particular provides further evidence that a hybrid synchronous approach using mechanistic structure with data-driven calibration can provide an ideal balance between prediction and interpretability. Indeed, as mentioned above, a more abstract mechanistic structure has extended the SEIR framework to include higher-level features such as seasonal forcing (time-varying transmission), contact heterogeneity, waning immunity and age structure[7,8]. All of these extensions can be thought of as providing an ideal balance that seeks to take advantage of both mechanistic knowledge of influenza transmission and the complex nature of real-world epidemics (i.e., filling gaps between theory and the ground-truth of on-the-ground epidemics).

There are many important, intrinsic limitations of mechanistic models, however. As described, these models are highly parameterized, and many of the model parameters (e.g., contact matrices, immunity waning, and vaccine uptake) are not directly observable and therefore are difficult to estimate with high confidence or reliability. Model misspecification or parameter nonidentifiability may also lead to limited prediction power, and it is well established that pure mechanistic models often do not outperform simple statistical or machine-learning models at short-term, real-time prediction.

3.2 Time-Series Forecasting (Data-Driven Models)

Statistical time-series models such as ARIMA/SARIMA and their covariate-enhanced variants (ARIMAX/SARIMAX) remain mainstays of influenza forecasting, primarily due to their relative conceptual simplicity, interpretability, and low data requirements. For example, analysis of ILI data from Chongqing, China (2010–2022) that exhibited a strong seasonal signal—with ILI case counts rising during the cold season (December–January) and showing a secondary peak in some years (May–June) among young children—has been conducted [9]. A SARIMAX model—i.e., SARIMA with external meteorological covariates—was fitted to generate short-term ILI forecasts, with model parameters pre-estimated via grid search based on AIC and post-validated using residual whiteness tests, and the results showed that the SARIMAX model had good short-term forecasting performance, supporting its potential for hospital-level operational planning [9].

Likewise, ILI data from Shanxi, China has been used to assess a hybrid SSA-SARIMA-LSTM model against control models including standalone SARIMA and SARIMA–LSTM hybrids [10]. The hybrid model outperformed single-method approaches across metrics including mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE), demonstrating that combining linear (SARIMA) and nonlinear (LSTM) components—along with signal decomposition via SSA—can yield higher predictive performance than standalone SARIMA. This case also illustrates that while SARIMA is often used as a benchmark for influenza forecasting, meaningful performance improvements can be achieved by coupling it with complementary analytical components.

Advantages of time-series methods include rapid fitting and updating as new weekly data becomes available, transparent structure with easily interpretable parameters, and strong short-term (1–2 week ahead) forecasting performance under stable seasonal regimes. Disadvantages of these methods are nevertheless notable. They are unable to explicitly model mechanistic dynamics, including intervention impacts and immunity waning, which limits their utility for scenario planning, such as predicting vaccine rollout effects. They are also sensitive to violations of assumptions, including nonstationarity from reporting changes and structural breaks from novel strains, leading to degraded performance during atypical epidemics, and short-term forecast accuracy decays rapidly beyond 1–2 weeks, especially if the epidemic curve deviates from historical seasonal profiles. As a result, while time-series methods can in theory be very useful for nowcasting or 1–2 week ahead forecasting, their extrapolative ability across seasons or with abrupt changes is limited. In practice, they often become a backbone of operational forecasting pipelines but are rarely sufficient in themselves for medium-range forecasts or scenario exploration.

3.3 Time Machine Learning and Ensemble Methods

Machine learning (ML) is explicitly designed to capture high-order complex nonlinear temporal patterns and associations among predictors, and of these approaches, LSTM (long short-term memory) networks are among the most popular choices for applied epidemiological time series forecasting problems. For instance, a hybrid LSTM + Genetic Algorithm (GA) model has been proposed to optimally tune hyperparameters, including network layers, units, and window size [11]. When applied to weekly U.S. ILI data sourced directly from CDC surveillance, this model significantly outperformed baseline statistical approaches and standard neural network models, particularly in multi-step ahead prediction. This outcome suggests ML models can appropriately handle the nonlinear dynamics of influenza spread and capture higher-order lag effects in time series when hyperparameters are carefully optimized.

Ensemble modeling approaches that pool predictions across multiple models can be particularly attractive for hedging against the weaknesses of individual models. For example, efforts to ensemble time-series, mechanistic, and ML model forecasts have attempted to address the variance in individual predictions while also improving the overall robustness of forecasts [5]. The previously described SSA-SARIMA-LSTM hybrid models can themselves be considered ensemble approaches, given that they are built from three separate components: signal decomposition (SSA), linear trend modeling (SARIMA), and nonlinear pattern capture (LSTM) [10]. These models have demonstrated forecasting success and further vindicate the proposition that portfolios of mutually complementary modeling approaches (as opposed to single-method strategies) are better suited to address the complex task of influenza forecasting[10,12].

Strengths and challenges are clustered in Table 3 according to three main themes. The strengths of these approaches include enhanced predictive skill over linear baselines (ML models can learn nonlinear and long-lag structure, e.g., LSTM and GA-tuned nets); robustness and stability from diversity (hybrids and ensembles fuse decomposition, linear and nonlinear signals); and actionability with interpretability (e.g., if a mechanistic scaffold is retained in the process). Other benefits include the a priori learning ability of ML models, allowing them to adapt to new patterns; the ability to capture complex interactions, lags, and contributions of covariates even in nonlinear relationships; and reliance on multiple models, which tends to be more stable and reliable than a single model.

The challenges of ML and ensemble approaches can be clustered around three main themes. Black-box models have limited interpretability, which impedes public health stakeholders from intuitively understanding a model's predictions. MLOps requirements are non-trivial and include hyperparameter tuning, revision-aware training and drift surveillance, along with careful hyperparameterization and retraining of ML models to prevent overfitting to past patterns. Ensemble governance can be a challenge: naïve averaging can be outperformed if an ensemble overweights an otherwise biased component model, and ensemble methods in general require weighting and calibration to prevent poor performance from a biased component model.

In sum, the empirical evidence suggests that hybrid, surveillance-aligned ensembles with online weighting and well-defined targets appear to be a prudent approach to balance forecast accuracy, stability, and transparency. In aggregate, ML and ensemble approaches are promising for medium-term influenza forecasts but demand careful validation, performance monitoring for drift and combination with more interpretable models to address their limitations in interpretability.

4. Synthesis and Suggestions

Synthesis of mechanistic and hybrid forecasting literature provides several actionable suggestions for practice to advance seasonal influenza forecasting in public health practice, and to bridge the gap between model development and real-world deployment. First, predictive frameworks should incorporate a hybrid modeling approach combining mechanistic modeling with data calibration. As evidenced by a SEIR-type model calibrated with separate compartments for vaccinated and unvaccinated populations (as well as a vaccine attenuation parameter), a mechanistic model can also be harnessed to directly estimate vaccine efficacy in near-real conditions [7]. This work suggests that mechanistic models should not be confined to purely academic exercises; by bridging a mechanistic base with surveillance data calibration that mimics real-world conditions, they can be leveraged as a principled framework for forecasting. Put differently, predictive systems should and can utilize hybrid calibration, applying a mechanistic structure in model formulation but fitting key parameters (e.g., transmission rate, vaccine attenuation) to calibration data in each influenza season. Second, hybrid and ensemble architectures should be prioritized for influenza forecasting. A hybrid LSTM and Genetic Algorithm model that optimizes hyperparameters for deep learning-based time series prediction was shown to outperform a baseline model for multi-step-ahead forecasts [11]. The success of this approach is promising and suggests that hybrid models combining statistical, mechanistic, and ML submodels should be used to capture the complementary strengths of each modality: statistical models for short-term stability, mechanistic models for interpretability, and ML for nonlinear pattern recognition. Ensemble methods aggregating forecasts across different models or approaches can further reduce variance and guard against overfitting to a single model structure. Third, ensemble weights and model parameters should be adaptively updated and not held constant across seasons. Model weights or parameters should be adjusted as new surveillance data becomes available (e.g., updates on case counts, subtype distribution, environmental conditions) to allow the forecasting system to adapt to new or novel outbreak features such as emergent virus strains, changes in population behavior (e.g., mobility shifts), or alterations to reporting practices. This update will be critical for maintaining forecast accuracy as epidemic conditions evolve. Fourth, forecasting targets and evaluation metrics should be standardized to facilitate translation to public health action. Forecasting systems should use canonical influenza targets, including epidemic onset week, peak week, and peak intensity, as well as scoring metrics compatible with established public health platforms (e.g., CDC FluSight) [5,13]. Standardized targets and metrics will allow for benchmarking of model performance across different regions, facilitate comparison of different forecasting approaches, and streamline integration of different models into ensemble systems. Finally, model documentation and drift monitoring should be operationalized to ensure long-term reproducibility.

5. Conclusion

Seasonal influenza is a predictable yet variable threat, whose understanding and anticipation can be improved by connecting drivers, surveillance hierarchies, and modeling approaches. This survey examines laboratory and epidemiologic evidence that absolute humidity modulates influenza persistence and transmissibility, underpinning expected winter seasonality in temperate zones. It describes the surveillance backbone from WHO FluNet/GISRS to national and regional dashboards (CDC FluView, ECDC ERVISS) that standardize weekly indicators and define forecasting targets, then contrasts three method families—statistical time-series (e.g., SARIMA/SARIMAX), mechanistic compartmental models (SEIR-type with seasonal forcing and vaccination), and machine-learning approaches (e.g., LSTM)—showing multi-model ensembles consistently outperform baselines for onset, peak week, and peak intensity. The forecasting pipeline is illustrated with case studies from the U.S., Europe, and China, highlighting short-term skill, value of climate-aware forcing, and operational challenges of data lags/revisions. These results support a hybrid, surveillance-aligned system that pairs an interpretable transmission structure with flexible data-driven calibration, ensemble weighting, and drift monitoring to enable earlier warnings and actionable scenario analysis. Operationalizing this approach as a compact toolkit can help public-health programs move from reactive response to proactive preparedness, informing timely allocation of vaccines, antivirals, and hospital resources. Limitations include imperfect surveillance data, context-dependent generalizability, model identifiability and drift, and early-stage MLOps for operational adoption; future work includes multi-pathogen data integration, adaptive ensembles, improved data engineering, and decision-theoretic evaluation.

References

[1] World Health Organization (WHO). “FluNet (Global Influenza Surveillance and Response System).” Available: https://www.who.int/tools/flunet. Accessed 1 Oct 2025.

[2] Centers for Disease Control and Prevention (CDC). “FluView—U.S. weekly influenza indicators and archives.” Available: https://www.cdc.gov/fluview/index.html Accessed 10 Oct 2025

[3] Broberg, E. K., Svartström, O., Riess, M., et al. (2024). Co-circulation of seasonal influenza A(H1N1)pdm09, A(H3N2) and B/Victoria lineage viruses with further genetic diversification, EU/EEA, 2022/23 influenza season. Euro Surveillance, 29(39): pii=2400020. https://doi.org/10.2807/1560-7917.ES.2024.29.39.2400020.

[4] Shaman, J., & Kohn, M. (2009). “Absolute humidity modulates influenza survival, transmission, and seasonality.” Proceedings of the National Academy of Sciences of the USA, 106(9), 3243–3248. https://doi.org/10.1073/pnas.0806852106.

[5] Reich, N. G., McGowan, C. J., et al. (2019). “Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S.” PLOS Computational Biology, 15(11), e1007486. https://doi.org/10.1371/journal.pcbi.1007486.

[6] Chen, X., Tao, F., Chen, Y., Cheng, J., Zhou, Y., & Wang, X. (2024; e-collection 2025). “Forecasting influenza epidemics in China using transmission dynamic model with absolute humidity.” Infectious Disease Modelling, 10(1), 50–59. https://doi.org/10.1016/j.idm.2024.08.003.

[7] Andreu-Vilarroig, C., Villanueva, R. J., & González-Parra, G. (2024). “Mathematical modeling for estimating influenza vaccine efficacy: A case study of the Valencian Community, Spain.” Infectious Disease Modelling, 9(3), 744–762. https://doi.org/10.1016/j.idm.2024.04.006.

[8] Hou, Y., & Bidkhori, H. (2024). “Multi-feature SEIR model for epidemic analysis and vaccine prioritization.” PLOS ONE, 19(3), e0298932. https://doi.org/10.1371/journal.pone.0298932.

[9] Chen, H., & Xiao, M. (2024). “Seasonality of influenza-like illness and short-term forecasting model in Chongqing from 2010 to 2022.” BMC Infectious Diseases, 24, 301. https://doi.org/10.1186/s12879-024-09301-4.

[10] Zhao, Z., Zhai, M., Li, G., Gao, X., Song, W., Wang, X., … Qiu, L. (2023). “Study on the prediction effect of a combined model of SARIMA and LSTM based on SSA for influenza in Shanxi Province, China.” BMC Infectious Diseases, 23(1), 71. https://doi.org/10.1186/s12879-023-08025-1.

[11] Kara, A. (2021). “Multi-step influenza outbreak forecasting using deep LSTM network and genetic algorithm.” Expert Systems with Applications, 180, 115153. https://doi.org/10.1016/j.eswa.2021.115153.

[12] Zhu, H., Chen, S., Lu, W., et al. (2022). Study on the influence of meteorological factors on influenza in different regions and predictions based on an LSTM algorithm. BMC Public Health, 22, 2335. https://doi.org/10.1186/s12889-022-14299-y.

[13] Reich, N. G., Brooks, L. C., Fox, S., Kandula, S., et al. (2019). “A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States.” Proceedings of the National Academy of Sciences of the USA, 116(8), 3146–3154. https://doi.org/10.1073/pnas.1812594116.