Contemporary Mathematics
http://ojs.wiserpub.com/index.php/CM/

Research Article

Bayesian Modeling of INGARCHX Models for Cellulitis Related to Meteorological Factors in Mahasarakham and Roi-Et Hospitals

Khemmanant K[1*], Kunwithree P[2], Sujitta S[1]

[1]Department of Mathematics, Faculty of Science, Mahasarakham University, Maha Sarakham 44150, Thailand
[2]Mathematics and Statistics Program, Kanchanaburi Rajabhat University, Kanchanaburi 71190, Thailand
E-mail: khemmanant.k@msu.ac.th

Received: 16 May 2024; **Revised:** 24 June 2024; **Accepted:** 9 July 2024

Abstract: The objective of this research is to develop integer-valued generalized autoregressive conditional heteroscedasticity (INGARCH) models to represent the weekly incidence of cellulitis cases in relation to two exogenous variables: seasonality and the weekly average of either relative humidity or maximum temperature. The key to the proposed model is its capacity to explain overdispersion and lag dependence. For predictions and model parameters, we employ the Bayesian Markov Chain Monte Carlo (MCMC) approach, as supported by both a simulation study and an empirical study. To assess different models, we apply the deviance information criterion (DIC) criterion to the weekly cellulitis case sample data with two independent variables. In addition, we offer a one-week prediction to help Mahasarakham and Roi-Et Hospitals manage the increasing volume of hospital admissions by estimating the weekly cellulitis case incidence rate.

***Keywords*:** cellulitis, INGARCH, posterior predictive distribution, overdispersion, MCMC, DIC, one-week-ahead

MSC: 62M10

1. Introduction

Bacterial infections of the subcutaneous fat layer and the lower layer of the dermis cause cellulitis. Patients present with fever, chills, and fatigue, along with a rash that is swollen, red, hot, unclear, and painful. The rash can spread wider, but not as quickly as with erysipelas. Possible complications include infections in the bloodstream, inflamed lymph nodes, glomerulonephritis, etc. In general, infection-related tissue disease occurs along the skin, even without any wounds. The causes of cellulitis come from two common types of bacteria: Streptococcus and Staphylococcus. Once infected, the skin will cause pain, heat, swelling, and redness, and then the area can spread quickly.

Factors that increase the risk of developing cellulitis include age, congenital diseases, occupational history, gender, and exposure to various environments. Variations in temperature can also affect the likelihood of infectious illness transmission. The incidence of cellulitis increases during the summer and fall in Australia. Cellulitis is most common in outpatient and emergency departments. According to the 2006 report on the incidence of cellulitis in the United States, there were approximately 2,500 cellulitis cases per 100,000 people per year, especially in men and the elderly. The death rate from cellulitis disease worldwide is approximately 0.7-1.8%. The rate of hospital admissions for patients

DOI: https://doi.org/10.37256/cm.5420244971

with cellulitis in the United States also increased with increasing environmental temperatures as the average monthly temperature increased [1]. Even short periods of elevated weather may increase the incidence of cellulitis [2]. The highest incidence of cellulitis overall occurs in subtropical and temperate climates [3]. There is a seasonality to the incidence of cellulitis, and a significant part of this seasonality appears to be related to increased temperatures [1].

Studies on the rates of people admitted to hospital outpatient and inpatient departments have drawn interesting attention. Numerous research endeavors employ diverse statistical methodologies to examine the prevalence of cellulitis. For instance, descriptive research in the outpatient department of Sai Noi Hospital revealed that there are 9.1% more patients with skin infections than non-infectious skin illnesses, with cellulitis accounting for 2.564% [4]. Cellulitis has a high correlation with the average monthly temperature. The incidence of cellulitis grew by 3.47% with rising temperatures, peaking in July-September and declining in December-February, according to the linear regression model [3]. Furthermore, hospitalized instances of cellulitis associated with rising temperatures are more likely to increase if the average temperature exceeds 90 °F, according to logistic regression models [5].

In public health service regions 7-10 in northeastern Thailand, the number of patients in subcutaneous infection outbreaks increased noticeably between 2014 and 2018. The majority of people work in the agricultural sector, which carries a significant risk of infection. Those with wounds or underlying medical conditions are more susceptible to infection during the rainy season [6, 7]. The statistical data indicates that during the rainy season, particularly in June and September, there is a tendency for the number of cellulitis patients seeking treatment at Roi-Et and Mahasarakham Hospital to increase. Cellulitis-related hospitalization rates are rising in Roi-Et and Mahasarakham Hospital, which is consistent with each province's elevated climatic factors. Therefore, we are interested in examining the cellulitis incidence in relation to two meteorological factors using time series analysis.

Numerous disciplines, including economics, epidemiology, and accident investigation, gather time series counts with copious amounts of data. Typically, they exhibit overdispersion and a non-normal distribution. We have run a number of statistical analyses on a chronologically sorted set of counts. Over the past ten years, modelers working with integer-valued GARCH (INGARCH) models have become interested in this overdispersion control technique. Ferland et al. [8] offers an INGARCH model with a Poisson distribution. Fokianos [9] provides an overview of significant advancements in the fields of counting and time series analysis. Zhu [10] gives either under- or over-dispersive count data with extended Poisson INGARCH models. To reduce overdispersion, Khamthong et al. [6] provides an INGRACHX model of the weekly incidences of necrotizing fasciitis.

Our goal in this work is to develop an INGARCH model that will account for the weekly incidence of cellulitis cases in relation to two exogenous variables: seasonality and relative humidity or temperature. With a modified Poisson distribution, the suggested model will be more adaptable for identifying data with a lower mean than variance. The first approach is more adaptable in terms of accounting for time-series counts. Furthermore, we offer a one-week head start, which facilitates prompt treatment planning to prevent death and significantly lower the prevalence of disability. We perform one-week prediction and parameter estimation for the suggested models using Bayesian approaches. Numerous studies have employed Bayesian techniques to effectively estimate parameters for time series of counts; they include Chen and Lee [11, 12], Chen et al. [13], Khamthong et al. [6].

The organization of this study runs as follows: Using a modified Poisson distribution, Section 2 implements the INGARCHX model in terms of two exogenous variables. For assessing the suggested model parameters, Section 3 presents the MCMC method and Bayesian inference techniques, as well as provides predictions, diagnostic tests, and model comparisons. Section 4 provides simulation studies to assess the suggested models. Section 5 presents an empirical study of weekly cellulitis cases from four datasets, with analytical findings derived from the models used. Section 6 contains concluding remarks.

2. INGARCHX models with two exogenous variables

Assume Y is a random variable with two parameters, ϕ and θ, specified over an integral value that is not negative. From Consul and Jain [14], we can define a generalized Poisson (GP) distribution as follows:

$$P(Y=y)=\begin{cases}\theta(\theta+\phi y)^{y-1}\dfrac{\exp(-\theta-\phi y)}{y!} & ,\ y=0,1,2,\ldots\\ 0 & ,\ \text{for}\ \ y>m\ \ \text{if}\ \ \phi<0,\end{cases}$$

where m (≥ 4) is the biggest positive integer, $\theta + \phi m > 0$ for $\phi < 0$, and $\max(-1, -\theta/m) \leq \phi < 1$ are the other variables. For the GP distribution, the variance and mean are:

$$Var(Y)=\frac{\theta}{(1-\phi)^3}\quad\text{and}\quad E(Y)=\frac{\theta}{1-\phi}.$$

When $\phi = 0$, that is, $E(Y) = Var(Y) = \theta$, the GP distribution transforms into the conventional Poisson distribution with mean θ. We consider that $\{Y_t\}$ denotes the time series of the weekly incidence of cellulitis cases and $\{X_{i,t}\}$, $i = 1, 2$ denotes two exogenous variables, where $X_{1,t}$ represents the weekly average of the highest temperature or relative humidity, and $X_{2,t}$ is the season period. We set the rainfall season to run from May to September, and alternate seasons are selected using the following dummy variables:

$$X_{2,t}=\begin{cases}1, & t\equiv\{\text{May, June, July, August, September}\}\mod(12)\\ 0, & \text{otherwise.}\end{cases}$$

Let us assume that the distribution of $Y_t|\mathcal{F}_{t-1}$, $\boldsymbol{X}_t$ is a GP distribution with two parameters, (θ, ϕ). Thus, the generalized Poisson distributive INGARCH model (GP-INGARCHX) is as follows:

$$Y_t\mid\mathcal{F}_{t-1},\boldsymbol{X}_t\sim\text{GP}(\theta_t^*,\phi),\ \ \theta_t^*=(1-\phi)\theta_t,$$

$$\theta_t=\alpha_0+\alpha_1Y_{t-1}+\beta_1\theta_{t-1}+\sum_{i=1}^{2}\gamma_iX_{i,t}. \tag{1}$$

The following are the usual conditions for the covariance, stationarity, and mean parameters:

$$\alpha_0>0,\ \ \alpha_1,\beta_1\geq 0,\ \ \alpha_1+\beta_1<1,\ \ \gamma_i\geq 0,\ \ i=1,2. \tag{2}$$

Likewise, the GP-INGARCHX model becomes a Poisson-distributive INGARCHX (P-INGARCHX) model when $\phi = 0$, as follows:

$$Y_t\mid\mathcal{F}_{t-1},\boldsymbol{X}_t\sim\text{Poi}(\theta_t),$$

$$\theta_t=\alpha_0+\alpha_1Y_{t-1}+\beta_1\theta_{t-1}+\sum_{i=1}^{2}\gamma_iX_{i,t}. \tag{3}$$

3. Bayesian inference

We apply the Bayesian method outlined below to obtain the model parameters for the INGARCHX models. Consequentially, the conditional probability function in Eq. (1) is as follows:

$$p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\vartheta}) = \prod_{t=2}^{n} \left\{ \frac{\theta_t (\theta_t + \phi Y_t)^{Y_t - 1}}{Y_t!} \exp[-(\theta_t + \phi Y_t)] \right\},$$

where θ_1 represents the output intensity and $\boldsymbol{\vartheta} = (\boldsymbol{\alpha}, \phi, \theta_1)'$ is the vector of all unknown parameters, with $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \beta_1, \gamma_1, \gamma_2)'$. We assume that the parameter ϕ has a uniform priority, where an indicator $I(A)$ is the set of ϕ satisfying the condition $0 \leq \phi < 1$. To give the necessary bounds on $\boldsymbol{\alpha}$, we use a uniform prior on the parameters $\boldsymbol{\alpha}$, which limits a constraint in Eq. (2)

$$p(\boldsymbol{\alpha}) \propto I(B),$$

where B satisfies Eq. (2)'s requirements. We observe that ϕ and θ_1 have priors that, respectively, follow beta and gamma distributions. In other words, $\phi \sim \text{Beta}(b_1, b_2)$ and $\theta_1 \sim \text{Gamma}(a_1, a_2)$, with the scale parameters being a_2 and b_2 and the shape parameters being a_1 and b_1.

For every group, the conditional posterior density is equal to the probability function times the prior density of that group, as follows:

$$p(\boldsymbol{\vartheta} \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\vartheta}_{\neq i}) \propto p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\vartheta}) p(\boldsymbol{\vartheta} \mid \boldsymbol{\vartheta}_{\neq i}),$$

with $\boldsymbol{\vartheta}_i$ standing for each parameter group, $p(\boldsymbol{\vartheta}_i)$ for its prior density, and $\boldsymbol{\vartheta}_{\neq i}$ for the vector containing all model parameters, excluding $\boldsymbol{\vartheta}_i$. We use the following parameter groups to draw inferences: (i) ϕ; (ii) θ_1; (iii) $\boldsymbol{\alpha}$, and assume that the parameter groups are priority independent.

Specifically, for the models provided in Eqs. (1) and (3), the conditional posterior distributions of ϕ, θ_1 and $\boldsymbol{\alpha}$ are as follows:

(i) The conditional posterior distribution for ϕ is:

$$p(\phi \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\alpha}, \theta_1) \propto p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\vartheta}) p(\phi).$$

(ii) The conditional posterior distribution for θ_1 is:

$$p(\theta_1 \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\alpha}, \phi) \propto p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\vartheta}) p(\theta_1).$$

(iii) The conditional posterior distribution for $\boldsymbol{\alpha}$ is:

$$p(\boldsymbol{\alpha} \mid \boldsymbol{Y}, \boldsymbol{X}, \phi, \theta_1) \propto p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\vartheta}) p(\boldsymbol{\alpha}).$$

Note: Eqs. (1) and (3) use the identical MCMC process for the INGARCHX model, with the exception of generating the ϕ in Eq. (3).

3.1 *Predicting*

We therefore utilize the present observation Y_t for one week ahead of the estimation for the purpose of anticipating $\{Y_{t+j}, j \geq 1\}$ by augmenting each MCMC sample iteration by one step. Upon exceeding the upper bound of the posterior predictive distribution of the 95% credible interval, Y_{t+1} yields the following prediction density for the model estimate for the next week:

$$p(Y_{t+1} \mid \boldsymbol{Y}, \boldsymbol{X}) = \int p(Y_{t+1} \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\vartheta}) p(\boldsymbol{\vartheta} \mid \boldsymbol{Y}, \boldsymbol{X}) d\boldsymbol{\vartheta}$$

$$\approx \frac{1}{N-M}\sum_{j=M+1}^{N} p(Y_{t+1} \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\vartheta}^{[j]}),$$

where the provided distribution is followed by the conditional predictive distribution $p(Y_{t+1}|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\vartheta}^{[i]})$, and N and M denote the sample sizes of total iterations and the burn-in, respectively.

3.2 *Diagnostic test and choosing model*

Following the fitting of several promising models, we consider the diagnostic and model selection. The deviance information criterion (DIC) combines complexity and goodness of fit metrics, both of which rely on deviance, to assess the relative fit of many competing models. Spiegelhalter et al. [15] defines the DIC as follows:

$$\text{DIC} = \bar{D} + p_D,$$

where $p_D = \bar{D} - D(\tilde{\boldsymbol{\vartheta}})$, $\tilde{\boldsymbol{\vartheta}}$ is the posterior mean of $\boldsymbol{\vartheta}$, and $\bar{D} + p_D = E_{\boldsymbol{\vartheta}|\boldsymbol{Y}}[D(\boldsymbol{\vartheta})]$. The MCMC procedure produces the DIC value as a byproduct. The smallest value DIC is the favorite among all competing models. We employ two of them to choose the best model among all the competing models.

We use the standardized residuals [16] to assess the fit of the proposed INGARCHX models in Eqs. (1) and (3):

$$\upsilon_t = \frac{Y_t - E(Y_t \mid \mathcal{F}_{t-1}, \boldsymbol{X}_t)}{\left[Var(Y_t \mid F_{t-1}, \boldsymbol{X}_t)\right]^{1/2}}, \ t = 1, \ldots, n.$$

For υ_t and υ_t^2, the mean and variance of these residuals with insignificant sequential correlation should be about 0 and 1, respectively. The model makes the assumption that the given model is reliable.

4. Study by simulation

We currently perform simulation studies of Bayesian estimation to assess the MCMC procedure's effectiveness. With a sample size of 500, we simulate the following two particular models:

$$\mathcal{M}_1 : Y_t \mid \mathcal{F}_{t-1}, \boldsymbol{X}_t \sim \text{GP}(\theta_t^*, \phi), \ \theta_t^* = (1-\phi)\theta_t,$$

$$\theta_t = 1 + 0.4Y_{t-1} + 0.2\theta_{t-1} + 0.2X_{1,t} + 0.3X_{2,t}$$

$$\mathcal{M}_2 : Y_t \mid \mathcal{F}_{t-1}, X_t \sim \text{Poi}(\theta_t),$$

$$\theta_t = 1 + 0.2Y_{t-1} + 0.2\theta_{t-1} + 0.2X_{1,t} + 0.3X_{2,t}$$

For both models $\mathcal{M}_1$ and $\mathcal{M}_2$, we randomly select the variable $X_{1,t}$ as an integer given a normal distribution $N(4, 1)$. Let $X_{1,t}$ represent the relative humidity or maximum temperature on a weekly average. We are now setting a dummy variable $X_{2,t}$, which is defined as follows:

$$X_{2,t} = \begin{cases} 1, & t \equiv \{5, 6, 7, 8, 9\} \mod (12) \\ 0, & \text{otherwise.} \end{cases}$$

We establish the following hyperparameters for the sensitivity analysis: the prior $\phi \sim \text{Beta}(20, 80)$, the prior θ_1, and

$\theta_1 \sim$ Gamma(5, 1). For models $\mathcal{M}_1$ and $\mathcal{M}_2$, we provide the following initial values: $\boldsymbol{\alpha} = (0.2, 0.1, 0.1, 0.1, 0.1)'$. Lastly, we choose the sample sizes for the burn-in and the whole iteration: $M = 8{,}000$ and $N = 30{,}000$, respectively. We simulate 500 replications and only save the fifth iteration of the sample period in order to make a choice about the proposed models.

Table 1. Simulation results obtained from 500 replications of models $\mathcal{M}_1$ and $\mathcal{M}_2$

Model	$\mathcal{M}_1$						$\mathcal{M}_2$					
Par.	True	Mean	Std	2.5%	97.5%	CP	True	Mean	Std	2.5%	97.5%	CP
α_0	1.0	1.1244	0.4224	0.3440	1.9651	0.978	1.0	1.0324	0.3646	0.3481	1.7454	0.974
α_1	0.4	0.3944	0.0434	0.3096	0.4795	0.980	0.2	0.1929	0.0437	0.1087	0.2798	0.952
β_1	0.2	0.1812	0.0798	0.0393	0.3438	0.990	0.2	0.1868	0.1060	0.0219	0.4161	0.988
γ_1	0.2	0.2015	0.0849	0.0493	0.3742	0.988	0.2	0.2069	0.0687	0.0773	0.3444	0.958
γ_2	0.3	0.3386	0.1615	0.0628	0.6814	0.976	0.3	0.3253	0.1316	0.0896	0.5993	0.962
ϕ	0.1	0.1016	0.0179	0.0685	0.1385	1.000						

Table 1 shows true values and simulation results. For models $\mathcal{M}_1$ and $\mathcal{M}_2$, these are the average posterior mean, standard error, and 95% Bayesian credibly intervals over 500 replications. The true values' coverage probabilities (CP) are the percentages of 95% posterior intervals produced by MCMC iterations. There is at least a 95% confidence level between the true values and the values covered by the posterior means of both models. The simulation results demonstrate the validity of the posteriors calculated using the suggested sampling strategy. Each of the 500 replications of the MCMC algorithm accurately discovers ϕ.

5. Analytical findings

In order to examine the INGARCHX models, we take into account the weekly time-series incidence of cellulitis cases from January 1, 2014, to December 31, 2020 (366 weeks in total) at Roi-Et Hospital and Mahasarakham Hospital. We consider two types of cellulitis patient records: outpatients (OPD) for Roi-Et Hospital, which stands for DS1, and inpatients (IPD) for Mahasarakham Hospital, which stands for DS2. Consider the time series plot of weekly counts and barcharts for datasets DS1 and DS2 in Table 2, as shown in Figures 1 and 2. Table 2 shows summary statistics for weekly cellulitis cases for two datasets, DS1 and DS2. It is evident that the variance of the DS1 and DS2 datasets is greater than their mean. As a result, DS1 and DS2 show signs of overdispersion.

Table 2. Weekly cellulitis case summary statistics

Dataset	Hospital	Type	Mean	Var	Min	Q_1	Median	Q_3	Max
DS1	Roi-Et	OPD	5.82	12.55	0	3	5	8	20
DS2	Mahasarakham	IPD	14.79	44.28	1	10	14	18.75	40

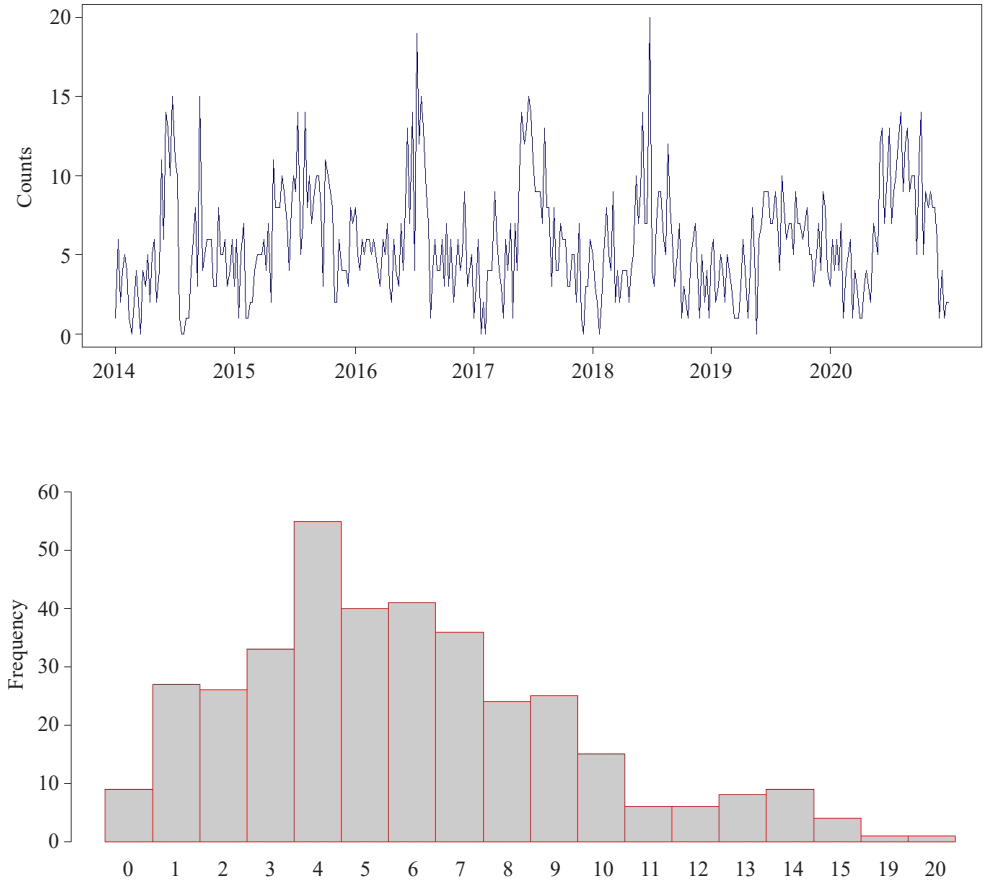
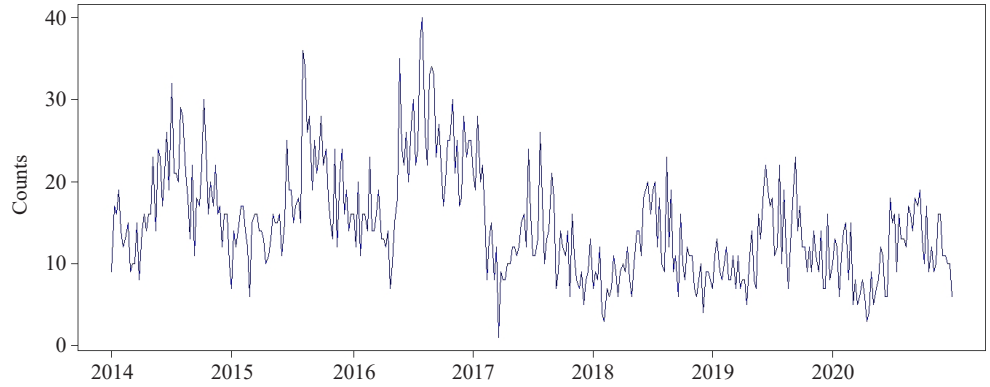


Figure 1. Time series: weekly cellulitis cases (top panel) and Barchart (bottom panel) for DS1



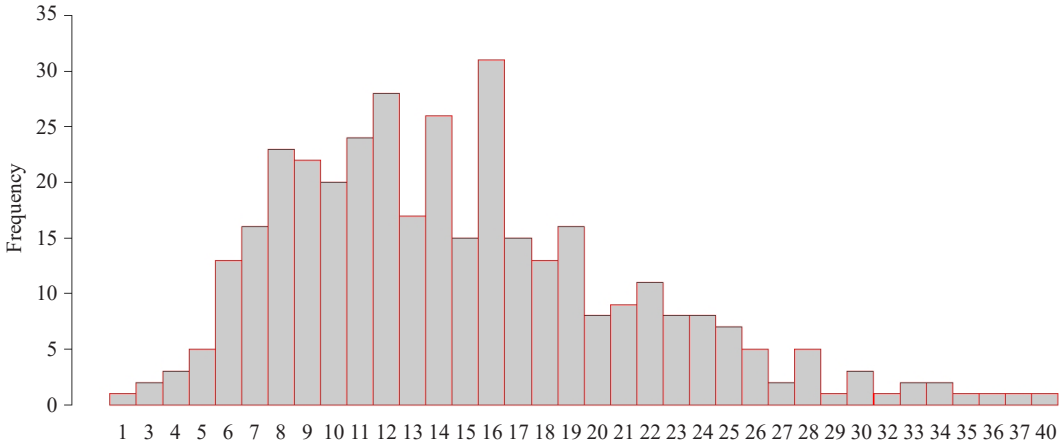


Figure 2. Time series: weekly cellulitis cases (top panel) and Barchart (bottom panel) for DS2

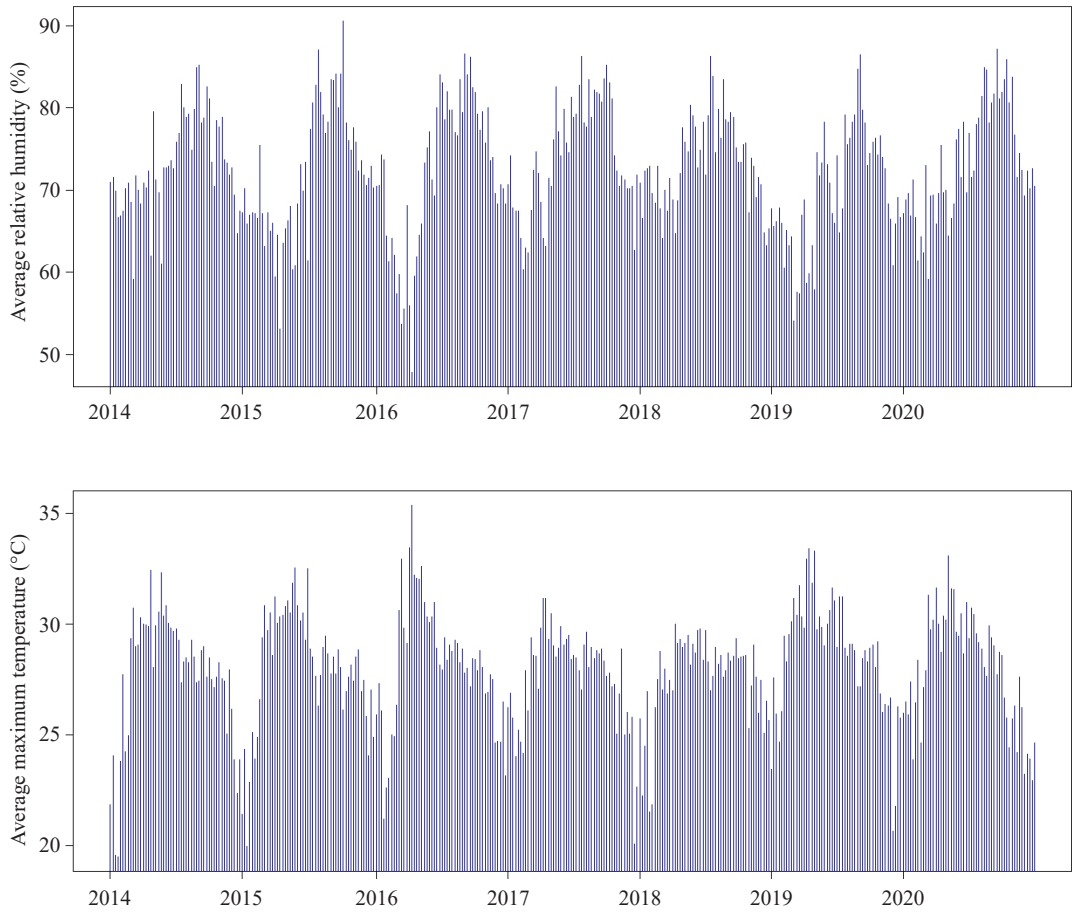


Figure 3. DS1: average maximum temperature and average relative humidity (top and bottom panels, respectively)

Figure 4. DS2: average maximum temperature and average relative humidity (top and bottom panels, respectively)

For two independent variables, we consider the meteorological data in Roi-Et and Maha Sarakham provinces obtained from the Meteorological Office. These are relative humidity, maximum temperatures, and seasonality. We now take relative humidity and maximum temperature as weekly averages, as shown in Table 3. Table 3 presents summary statistics for the weekly average maximum temperature and the weekly average relative humidity. These two independent variables can plot time series as shown in Figures 3-4.

Table 3. Weekly meteorological variable summary statistics

Province	Dataset	Mean	Median	Min	Max
Roi Et	Average relative humidity (%)	72.10	71.10	44.57	91.57
	Average maximum temperature (°C)	27.85	28.25	19.71	35.14
Maha Sarakham	Average relative humidity (%)	72.56	72.43	47.86	90.57
	Average maximum temperature (°C)	27.98	28.43	19.49	35.37

Dummy variables are in the seasonal data set for both the rainy (May-September) and other seasons. This is the seasonal variable, $X_{2,t}$. We set $X_{2,t}$ to be a dummy variable with a value of 1 and 0 for the rainy season and otherwise, respectively, as follows:

$$X_{2,t} = \begin{cases} 1, & \equiv \{5,6,7,8,9\} \mod(12) \\ 0, & \text{Otherwise.} \end{cases}$$

In practice, we divided the cellulitis cases in Rot-Et Hospital and Mahasarakham Hospital into four datasets: DS1-1, DS1-2, DS2-1, and DS2-2, as given in Table 4.

Table 4. Classification for four weekly datasets

Hospital	Dataset	Classification
Roi-Et	DS1-1	The weekly cellulitis cases include the weekly mean relative humidity and seasonality.
	DS1-2	The weekly cellulitis cases include the weekly maximum mean temperature and seasonality.
Mahasarakham	DS2-1	The weekly cellulitis cases include the weekly mean relative humidity and seasonality.
	DS2-2	The weekly cellulitis cases include the weekly maximum mean temperature and seasonality.

These two weekly independent variables, X_1 and X_2, are not multicollinear, which indicates that all two variables can help describe weekly cellulitis cases, as shown in Tables 5 and 6. Therefore, to model cellulitis cases in Rot-Et and Mahasarakham Hospital, we can simultaneously use X_1 and X_2, and we can also see that each dataset of two variables correlates independently. For convenience in evaluating the model parameters in equations (1) and (3), we transform either the weekly average relative humidity variable or the weekly average maximum temperature variable, $X_{1,t}$, to be:

$$X_{1,t}^{*} = \frac{X_{1,t} - \min(X_{1,t})}{S_{XX}^{(1)}}, \quad t = 1, \ldots, n$$

where $S_{XX}^{(1)}$ is used in place of the standard deviation of $X_{1,t}$.

Table 5. Correlation matrices of variables for DS1-1 and DS2-1

Province	Roi-Et			Maha Sarakham		
Variable	Y	X_1	X_2	Y	X_1	X_2
Y	1	0.40	0.14	1	0.34	0.20
X_1	0.40	1	0.44	0.34	1	0.43
X_2	0.14	0.44	1	0.20	0.43	1

Y: DS1-1 (DS2-1); X_1: average relative humidity; X_2: seasonal

Table 6. Correlation matrices of variables for DS1-2 and DS2-2

Province	Roi-Et			Maha Sarakham		
Variable	Y	X_1	X_2	Y	X_1	X_2
Y	1	0.21	0.15	1	0.06	0.20
X_1	0.21	1	-0.11	0.06	1	-0.11
X_2	0.15	-0.11	1	0.20	-0.11	1

Y: DS1-2 (DS2-2); X_1: average relative humidity; X_2: seasonal

Table 7. Bayesian parameter estimations INGARCHX models for DS1-1

Model	GP-INGARCHX					P-INGARCHX				
Par.	Mean	Median	Std	2.5%	97.5%	Mean	Median	Std	2.5%	97.5%
α_0	0.6457	0.6369	0.3574	0.0411	1.3776	0.7016	0.6913	0.3594	0.0733	1.3966
α_1	0.2310	0.2298	0.0412	0.1528	0.3157	0.2631	0.2625	0.0399	0.1860	0.3392
β_1	0.3930	0.3970	0.0883	0.2050	0.5558	0.4281	0.4328	0.0727	0.2771	0.5595
γ_1	0.0754	0.0685	0.0496	0.0037	0.1817	0.0812	0.0763	0.0530	0.0046	0.1915
γ_2	0.9689	0.9529	0.2384	0.5330	1.4598	1.1474	1.1330	0.2326	0.7374	1.6412
ϕ	0.1625	0.1618	0.0243	0.1148	0.2122					
DIC	-3,534.3025					-3,522.1964				

Table 8. Bayesian parameter estimations INGARCHX models for DS1-2

Model	GP-INGARCHX					P-INGARCHX				
Par.	Mean	Median	Std	2.5%	97.5%	Mean	Median	Std	2.5%	97.5%
α_0	0.6153	0.6037	0.3367	0.0554	1.3018	0.7113	0.7213	0.3739	0.0533	1.4432
α_1	0.2305	0.2295	0.0423	0.1514	0.3173	0.2647	0.2641	0.0422	0.1825	0.3473
β_1	0.4183	0.4213	0.0829	0.2536	0.5677	0.4448	0.4490	0.0704	0.3015	0.5743
γ_1	0.0461	0.0436	0.0292	0.0023	0.1092	0.0481	0.0442	0.0319	0.0021	0.1140
γ_2	0.9540	0.9389	0.2307	0.5372	1.4230	1.1405	1.1234	0.2314	0.7105	1.6293
ϕ	0.1634	0.1635	0.0244	0.1160	0.2124					
DIC	-3,534.4578					-3,520.8808				

Table 9. Bayesian parameter estimations INGARCHX models for DS2-1

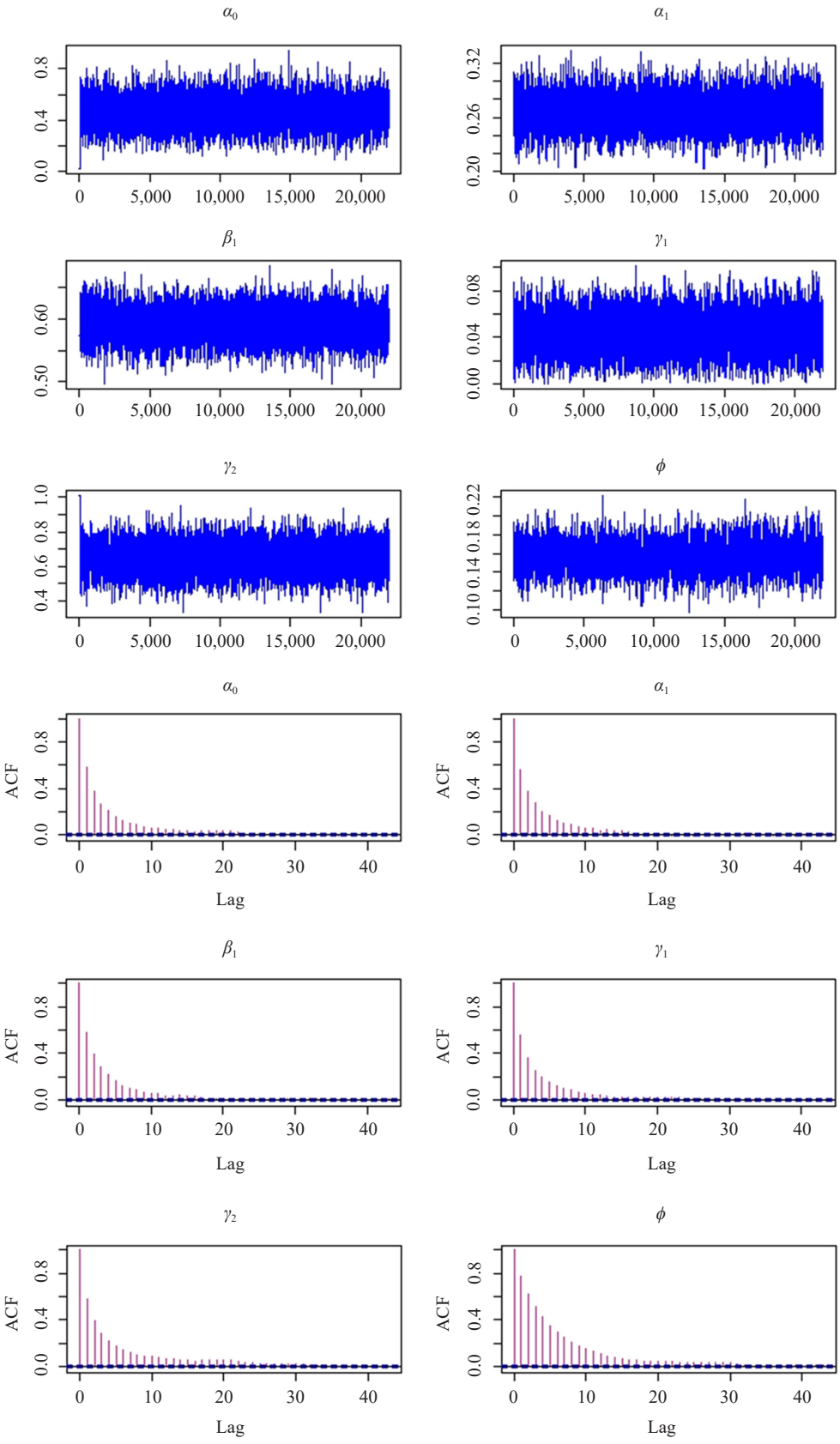
Model	GP-INGARCHX					P-INGARCHX				
Par.	Mean	Median	Std	2.5%	97.5%	Mean	Median	Std	2.5%	97.5%
α_0	0.4510	0.4537	0.1208	0.2121	0.6852	0.6229	0.6234	0.1797	0.2706	0.9799
α_1	0.2633	0.2636	0.0183	0.2283	0.2984	0.3037	0.3032	0.0220	0.2626	0.3487
β_1	0.5976	0.5977	0.0247	0.5496	0.6455	0.6097	0.6101	0.0289	0.5513	0.6646
γ_1	0.0414	0.0410	0.0163	0.0099	0.0743	0.0291	0.0288	0.0126	0.0060	0.0547
γ_2	0.6658	0.6633	0.0867	0.4961	0.8444	0.8989	0.8980	0.0985	0.7068	1.0897
ϕ	0.1562	0.1562	0.0156	0.1256	0.1867					
DIC			-18,915.1266					-18,905.2867		

Table 10. Bayesian parameter estimations INGARCHX models for DS2-2

Model	GP-INGARCHX					P-INGARCHX				
Par.	Mean	Median	Std	2.5%	97.5%	Mean	Median	Std	2.5%	97.5%
α_0	0.2473	0.2467	0.0960	0.0629	0.4382	0.3700	0.3708	0.1089	0.1549	0.5780
α_1	0.2618	0.2618	0.0202	0.2222	0.3019	0.2921	0.2918	0.0218	0.2481	0.3350
β_1	0.6077	0.6078	0.0278	0.5522	0.6620	0.6285	0.6289	0.0258	0.5771	0.6784
γ_1	0.0451	0.0453	0.0123	0.0213	0.0690	0.0377	0.0373	0.0152	0.0090	0.0679
γ_2	0.7111	0.7109	0.1173	0.4849	0.9490	0.9555	0.9578	0.1203	0.7155	1.1894
ϕ	0.1570	0.1567	0.0161	0.1262	0.1886					
DIC			-18,913.0231					-18,908.2205		

Tables 7-10 provide Bayesian estimates of INGARCHX models for datasets: DS1-1, DS1-2, DS2-1, and DS2-2. These Bayesian estimates consist of posterior means, medians, standard errors, and 95% credible intervals. All Bayesian parameter estimates for each dataset correspond to the offered models.

To monitor the convergence of all parameters of the fitted GP-INGARCHX model, we present two plots: trace and ACF plots, as shown in Figures 5-6. Overall, the evidence supports the convergence of the MCMC method. The outcomes are satisfying, and the convergence is clear for each dataset (DS1-1, DS1-2, DS2-1, and DS2-2), as demonstrated through ACF plots. The ACF plots reduce the autocorrelations on MCMC iterations.



DS1-1

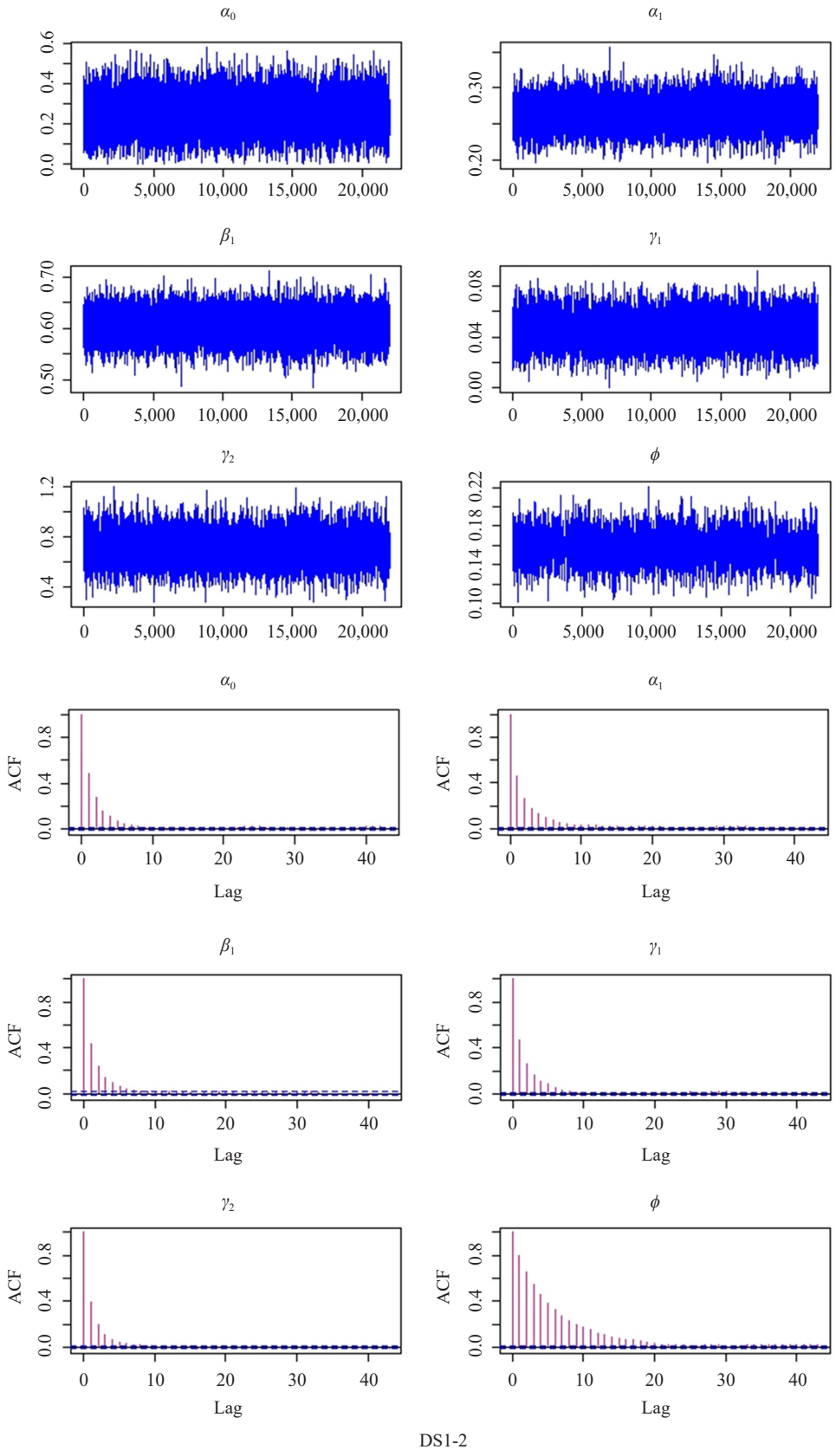
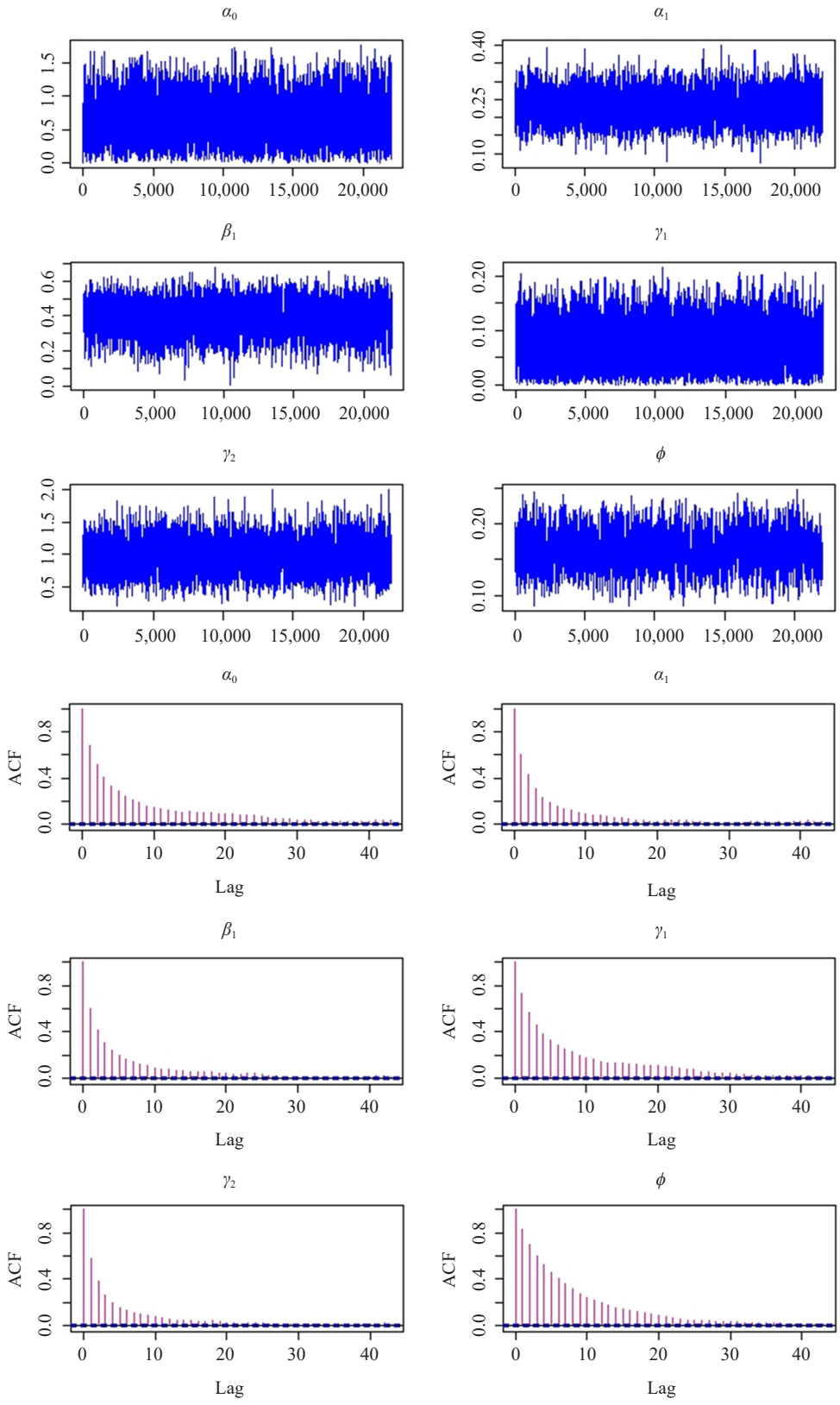


Figure 5. Trace and ACF plots of the MCMC estimates DS1-1 and DS1-2 in the GP-INGARCHX model


α_0
α_1
β_1
γ_1
γ_2
ϕ
0.0 0.5 1.0 1.5
0.10 0.25 0.40
0.0 0.2 0.4 0.6
0.00 0.10 0.20
0.5 1.0 1.5 2.0
0.10 0.20
0 5,000 10,000 15,000 20,000
ACF
0.0 0.4 0.8
0 10 20 30 40
Lag

DS2-1

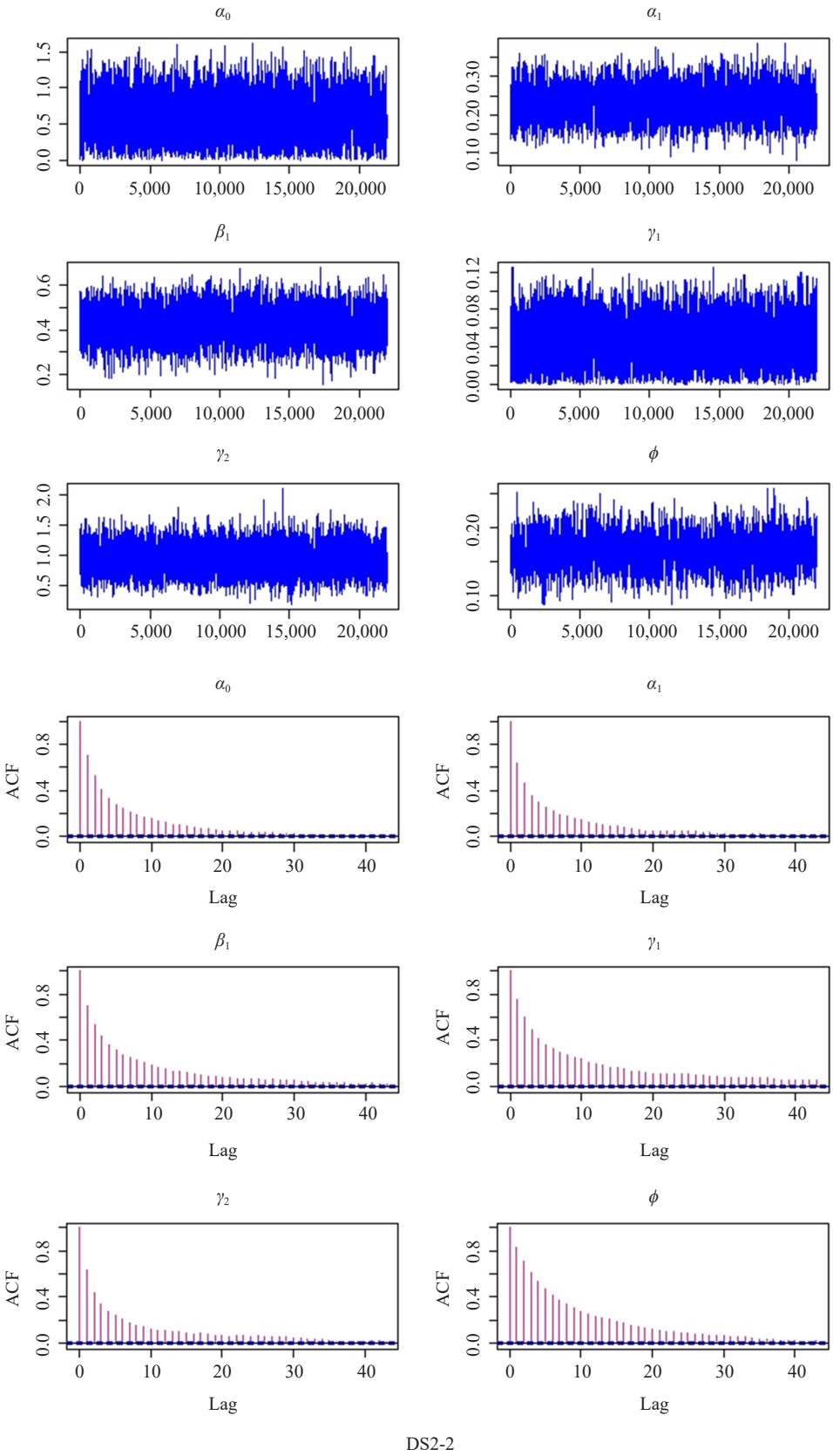


Figure 6. Trace and ACF plots of the MCMC estimates DS2-1 and DS2-2 in the GP-INGARCHX model

We take into consideration two models proposed in equations (1) and (3) for model comparison. According to the last row in Tables 7-10, the model that fits the data the best among the two competing models for DS1-1, DS1-2, DS2-1, and DS2-2 is the one with the lowest DIC values. All four datasets favor the model GP-INGARCHX.

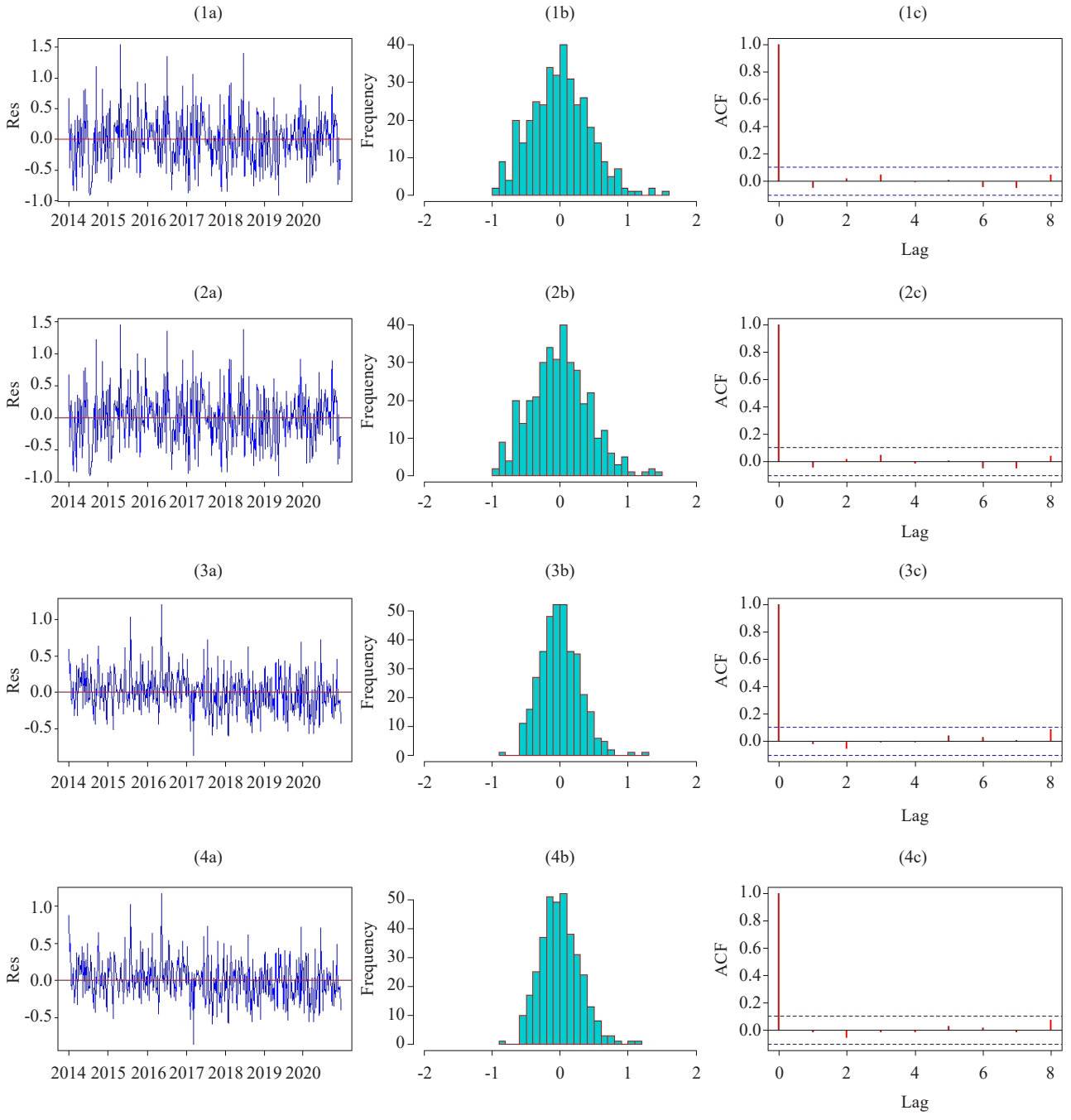


Figure 7. The study employed four datasets to perform diagnostic tests using standardized residuals on the selected models. These datasets included time series (1a-4a), histograms (1b-4b), and ACF plots (1c-4c) for DS1-1, DS1-2, DS2-1, and DS2-2, respectively

We use the standardized residuals in the target model's posterior estimations to test the diagnostic model's appropriateness using the Ljung-Box statistic. To assess the standardized residuals from Lags 1 to 20, use the Ljung-

Box statistics. The residuals indicate that, as Table 11 shows, the squared series is unable to reject the fitted model at the 0.05 significant level. Figure 7 displays the autocorrelation plots (1a)-(4a) for the standard residual time series, (1b)-(4b) for the standard residual histogram, and (1c)-(4c) for every best model. The residuals exhibit near-zero dispersion.

To predict the incidence one week ahead for all datasets, we use the best-fitted model. In Figures 8-9, the blue line shows the upper bound of the 95% credible interval for the prediction of all datasets, and the vertical solid lines represent Y_{t+1}. The red triangles represent the observed counts that are greater than the upper bound of prediction for all DS1-1, DS1-2, DS2-1, and DS2-2, which reflect the forecasting. We can observe that the red triangles specify at most 2 weeks of differences for DS1 and DS2. That is, the rate of exceeding the upper bound probability of 95% of datasets, DS1-1 (DS2-1) and DS1-2 (DS2-2), is almost equal.

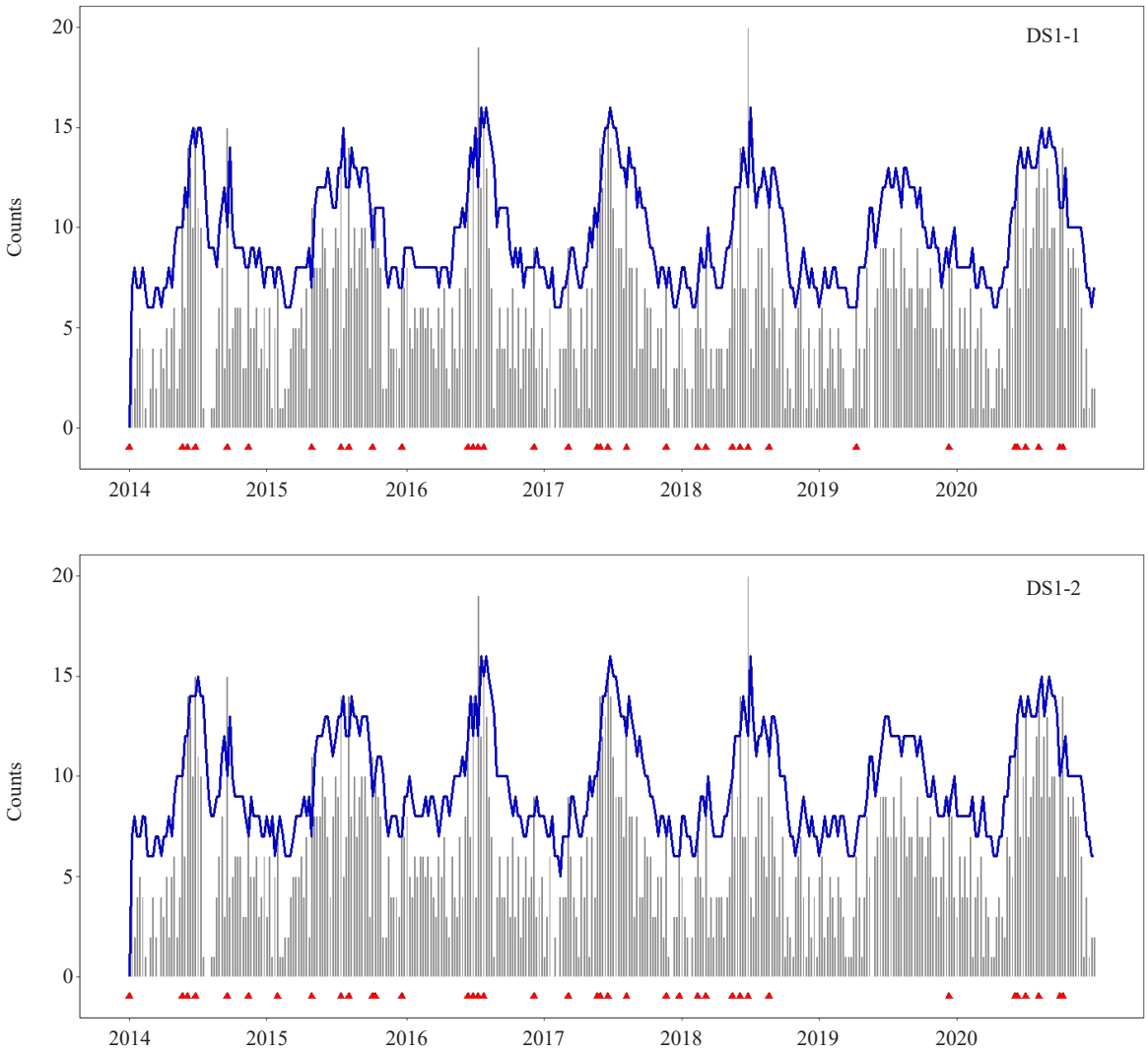


Figure 8. Utilizing the GP-INGARCHX model, DS1-1 and DS1-2 forecast one week in advance. The red triangles on DS1-1 and DS1-2 show that the observed numbers are higher than the upper bound of the 95% Bayesian credible intervals, which is in line with the prediction. The blue line shows the predictions for datasets DS1-1 and DS1-2. The solid line refers to Y_{t+1}

Table 11. Ljung-Box Q statistics of the standarized residuals based on the best-fitting models

Dataset	DS1-1		DS1-2		DS2-1		DS2-2	
Lag	Q	p-value	Q	p-value	Q	p-value	Q	p-value
1	0.738	0.390	0.625	0.429	0.110	0.739	0.043	0.836
4	1.728	0.786	1.554	0.817	1.066	0.900	1.036	0.904
8	3.849	0.871	3.793	0.875	4.861	0.772	3.620	0.900
12	14.058	0.297	14.137	0.292	6.760	0.873	5.436	0.942
16	17.036	0.383	16.911	0.391	10.875	0.817	8.912	0.917
20	20.351	0.436	20.676	0.416	17.015	0.652	15.144	0.768

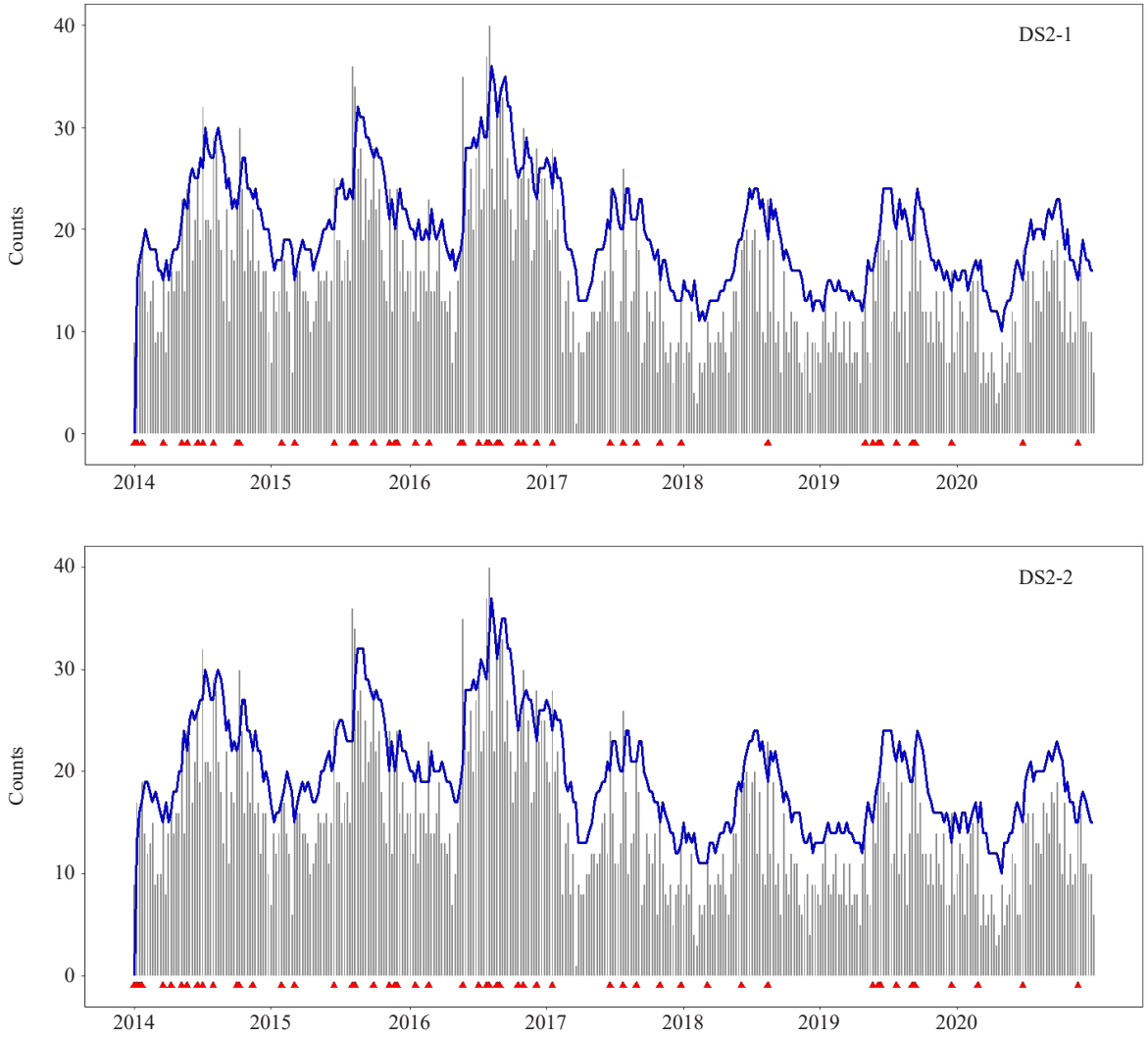


Figure 9. Utilizing the GP-INGARCHX model, DS2-1 and DS2-2 forecast one week in advance. The red triangles on DS2-1 and DS2-2 show that the observed numbers are higher than the upper bound of the 95% Bayesian credible intervals, which is in line with the prediction. The blue line shows the predictions for datasets DS2-1 and DS2-2. The solid line refers to Y_{t+1}

Our findings indicate that the incidence of weekly cellulitis patients in the outpatient department of Roi-Et Hospital and in the inpatient department of Mahasarakham Hospital related to weather variables tends to increase significantly. The rate of exceedance probability of the upper bound of prediction is 10.45% and 11.05% for DS1-1 and DS1-2, and 9.86% and 10.14% for DS2-1 and DS2-2. The majority is between May and September (rainy season), representing at least 5.97% and 5.48% for Roi-Et and Mahasarakham Hospital, respectively. This reseacrh points out that the predictions of the incidence of weekly cellulitis cases via the GP-INGARCHX model are significantly close. Hence, climate conditions such as relative humidity, temperature, and season have a significant effect on the increased incidence of cellulitis disease, which is consistent with patient data coming to receive services at both Roi-Et and Mahasarakham Hospitals.

6. Conclusion

This study proposes an INGARCH model for weekly cellulitis associated with climale variables such as seasonality, the weekly average maximum temperature, and relative humidity. The proposed model manages overdispersion and lagging dependencies, which is a crucial aspect. For modeling, inference, and prediction, we use Bayesian Markov Chain Monte Carlo (MCMC) approaches. The simulation results generated with the proposed MCMC approach provide accurate estimates and conclusions for each model parameter in the INGARCHX models. Furthermore, we compare our results with those of the two competing models using anticipated higher thresholds and model credibility intervals. Using a generalized Poisson, we are able to effectively use the INGARCHX model to describe the features of the weekly incidence of cellulitis and two meteorological variables across all datasets.

Our research reveals that two exogenous factorstemperature, relative humidity, and seasonhave a major impact on the high incidence of cellulitis. These findings align with the patient data from the Mahasarakham and Roi-Et hospitals. Predicting the probability that the counts would recur based on available data is one technique to help explain the features and evaluate the time series of the counts that needs more research. For count data, interest modeling takes into account time series data with various distributions (e.g., Poisson and a negative binomial) on a daily, weekly, monthly, and annual basis.

In order to characterize various statistical properties, the researcher can choose from a variety of INGARCHX modeling variations that use different distributions to account for the dataset under examination, whether it is linear or nonlinear. Furthermore, the number of variables in the model may rise or fall in accordance with the validity and usefulness of the count data.

Acknowledgements

This research project was financially by Mahasarakham University. We would like to thank Mahasarakham and Roi-Et Hospital for the information on patients with cellulitis, as well as the data on climate in each metrological station collected from the Thai Meteorological Department (TMD).

Conflict of interest

The authors declare no competing interest.

References

[1] Peterson RA, Polgreen LA, Cavanaugh JE, Polgreen PM. Increasing incidence, cost, and seasonality in patients hospitalized for cellulitis. *Open Forum Infectious Diseases*. 2017; 4(1): ofx008.

[2] Polgreen PM, Polgreen EL. Infectious diseases, weather, and climate. *Clinical Infectious Diseases*. 2017; 66: 815-817.

[3] Hsu RJ, Chou CC, Liu JM, Pang ST, Lin CY, Chuang HC, et al. The association of cellulitis incidence and meteorological factors in Taiwan. *Epidemiology and Infection*. 2019; 147: 1-4.
[4] Phacharawichit P. Prevalence of skin diseases in Sai Noi hospital. *Medical Journal of Regions*. 2020; 39: 440-453.
[5] Peterson RA, Polgreen LA, Sewell DK, Polgreen PM. Warmer weather as a risk factor for cellulitis: A population-based investigation. *Clinical Infectious Diseases*. 2017; 65: 1167-1173.
[6] Khamthong K, Srisarakham N, Suraphee S. Bayesian time series modeling of necrotizing fasciitis count in Mahasarakham and Roi-Et hospitals. *Lobachevskii Jounal of Mathematics*. 2023; 44(9): 3717-3727.
[7] Khamthong K, Phamrung K, Suraphee S. Bayesian modeling for the count of necrotizing fasciitis related to cellulitis and climate in Mahasarakham and Roi-Et hospitals. *Peace and Conflict*. 2024; 30(1): 58-74.
[8] Ferland R, Latour A, Oraichi D. Integer-valued GARCH processes. *Journal of Time Series Analysis*. 2006; 27: 923-942.
[9] Fokianos K. Count time series models. In Subba Rao T, Subba Rao S, Rao CR. (eds). *Handbook of Statistics 30: Time Series Analysis: Methods and Applications*. North Holland; 2012. p.315-347.
[10] Zhu F. Modeling overdispersed or underdispersed count data with generalized Poisson integervalued GARCH models. *Journal of Mathematical Analysis and Applications*. 2012; 389: 58-71.
[11] Chen CWS, Lee S. Generalized poisson autoregressive models for time series of counts. *Computational Statistics and Data Analysis*. 2016; 99: 51-67.
[12] Chen CWS, Lee S. Bayesian causality test for integer-valued time series models with applications to climate and crime data. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 2017; 66: 797-814.
[13] Chen CWS, Khamthong K, Lee S. Markov switching integer-valued generalized autoregressive conditional heteroscedastic models for dengue counts. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 2019; 68: 963-983.
[14] Consul PC, Jain GC. A generalization of the Poisson distribution. *Technometrics.* 1973; 15: 791-799.
[15] Spiegelhalter D, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B: Mathematics*. 2002; 64: 583-616.
[16] Jung RC, Kukuk M, Liesenfeld R. Time series of count data: modelling and estimation and diagnostics. *Computational Statistics and Data Analysis*. 2006; 51: 2350-2364.