

**15th National Convention on Statistics**  
Crowne Plaza Manila Galleria, Ortigas Center, Quezon City  
October 3-5, 2022

## **BAYESIAN LOG-LINEAR POISSON AUTOREGRESSIVE MODEL**

*by*

**Al-Ahmadgaid B. Asaad**

For additional information, please contact:

Author's name	Al-Ahmadgaid B. Asaad
Designation	Senior Data Scientist
Affiliation	Aboitiz Data Innovation PTE. LTD.
Address	Goldbell Towers, 47 Scotts Road, Singapore City, 228233, Singapore
Email	alahmadgaid@gmail.com
Website	<a href="https://al-asaad.github.io/">https://al-asaad.github.io/</a>

# Bayesian Log-Linear Poisson Autoregressive Model

Al-Ahmadgaid B. Asaad

## Abstract

The study extends the log-linear Poisson autoregressive model proposed by Fokianos and Tjøstheim (2011) into a Bayesian framework, and applies it to modeling the yearly tropical cyclone frequencies within the Philippine Area of Responsibility. The data, taken from the Japan Meteorological Agency's Regional Specialized Meteorological Center, with time points running from 1950 to 2021, is modeled at a semester-level granularity to take advantage of the periodic patterns suitable for autoregressive modeling. The estimation is done using Markov Chain Monte Carlo algorithm via the No-U-Turn Sampler. Finally, the proposed model can infer the characteristics of the Philippine storm frequencies, both in terms of the point forecasts and the associated uncertainties.

**Keywords:** Count Time Series, Julia, Markov Chain Monte Carlo, No-U-Turn Sampler, Typhoons

## 1 Introduction

With a maximum sustained wind speed of 315 kilometers per hour (kph) or 170 knots, and gusts reaching 379 kph (205 knots) just before landfall (Lagmay et al., 2015), super typhoon Haiyan—known locally as Yolanda—became one of the most powerful hurricanes ever recorded in history. It left the Philippines with 6300 deaths, 1061 missings and 28689 injured (Lagmay et al., 2015). Such disaster forced coconut farmers to explore other sources of income, including raising livestock, growing other crops, etc. after Yolanda reduced their yearly income by more than 400 USD/hectare (Seriño et al., 2021). These statistics are the very reasons why it is important to study the characteristics of the tropical cyclones (TCs), especially for the Philippines.

Among the efforts to understanding these characteristics is the study of the annual frequencies of the TCs, where the aim is to come up with a model that can accurately forecast future storm activities, and be used as part of the early warning systems for policy makers. Early works for the Western North Pacific (WNP) region, where the Philippines is located, focused on finding for independent covariates that can be used as predictors for modeling the TC frequencies. These covariates are atmospheric/oceanic signals such as the El Niño/Southern Oscillation (ENSO) phenomenon, where indices known to affect this event like the sea-surface temperature (SST) were studied by Chan (1985), whose findings motivated Chan (1995) to also study indices for the quasi-Biennial Oscillation (QBO) as possible predictors for the TC activities in the WNP.

Using the findings mentioned above, Chan et al. (1998) finally developed a statistical model for WNP TC activities using the following predictors: *i.* ENSO indices; *ii.* the environmental conditions over East Asia and the WNP from April of the previous year to March of the current year; and, *iii.* climatology and persistence (see Murphy (1992)) predictors. The model turn out to do quite well despite the findings of Lander (1994) that the observed annual TC totals in the WNP region are not significantly affected by any ENSO indices. The model used by Chan et al. (1998) was a generalization of Projection Pursuit Regression (PPR) named Smooth Multiple Additive Regression Technique (Friedman, 1985), which was further improved in terms of predictors by Chan et al. (2001). It should be noted that, Wang and Chan (2002) showed how strong the effect of ENSO on the TC activities in the WNP region, addressing earlier findings of Lander (1994). The quest to look for other indices, driven Fan (2007) to use a North Pacific Oscillation (NPO) index as part of the covariates using multiple linear regression model as the prediction equation. Another related work is the study by Choi et al. (2010) on the use of three teleconnection patterns: the Siberian High Oscillation (SHO) index in the East Asian continent; the NPO; and, the Antarctic Oscillation (AAO) near the Australia during the boreal spring (April-May). The indices found for these three teleconnection patterns were used

as predictors for their multiple linear regression model. Finally, Fan and Wang (2009) proposed a new approach to modeling TC activities, where instead of modeling the yearly totals, they proposed to model the year-over-year change of the frequency instead. The model used was also a multiple linear regression model.

Statistically, TC frequencies can be assumed to be Poisson distributed, so that later papers have incorporated this to their modeling. For example, Zhang et al. (2018) developed a Poisson regression model to predict the typhoon genesis formation in the WNP using April SST anomalies in the North Pacific region and SST anomalies over the Coral Sea off east coast of Australia. McDonnell and Holbrook (2004a) also used Poisson regression with saturated equivalent potential temperature gradient as part of its predictors for the Australian region (see also McDonnell and Holbrook (2004b)). Moreover, recent papers such as Magee et al. (2021), used multivariate Poisson regression to develop a location-specific TC frequency outlook using 16 covariates, among of which is the Trans NINO Index (TNI)—difference in normalized SST anomalies between the NINO1+2 ( $0^{\circ}$ – $10^{\circ}$ S,  $90^{\circ}$ – $80^{\circ}$ W) and NINO4 ( $5^{\circ}$ N– $5^{\circ}$ S,  $160^{\circ}$ E– $150^{\circ}$ W) regions. The locations targeted by the said paper are countries in the South East Asia region, including the Philippines.

Combinations of dynamical and statistical modeling have been explored as well, some of the works include Li et al. (2013) who combined the National Centers for Environmental Prediction (NCEP) Climate Forecast System version 2 (CFSv2)—a fully coupled ocean-land-atmosphere dynamical model—with linear regression model. The idea of a hybrid model is that, the forecast of the dynamical model is used as predictor for the statistical model. As such, Zhang et al. (2017) used dynamical model forecasts of the GFDL Forecast-Oriented Low Ocean Resolution version of CM2.5 with flux adjustments (FLOR-FA) as input for their Poisson regression. Finally, Tian and Fan (2019) uses the approach done by Fan and Wang (2009) but using a hybrid model.

While Poisson log-linear regression modeling does appropriately capture or at least approximate the true variations of the TC counts, the past values of it should also be taken into our advantage. Indeed, in time series modeling, the core motivation is that, variations of the response variable are expected to be time dependent. As for climatic events, the fact that dynamics are continuous over time, from gradual changes in temperature to movements in atmospheric circulations, simply suggest that, by nature, current weather patterns can be described by its preceding state. However, when these continuous changes are aggregated to discrete time points summarized across a wide temporal horizon like annual, a lot of informations are loss, making it difficult to model the variations, and hence most literatures resort to finding sets of independent covariates. This issue is of particular interest in this study. Among the literatures mentioned, Tian and Fan (2019) addressed the above limitation by incorporating a autoregressive term albeit not for a Poisson regression, since they modeled the yearly change of the frequency count.

Early works on the use of Autoregressive (AR) model (see Chapter 8 of Hyndman and Athanassopoulos (2018)) for count time series modeling was seen as problematic, mainly due to the “*multiplication problem*” (Weiss, 2018). To address this, McKenzie (1985) proposed some solutions which motivated the interger-valued AR(1) (INAR(1)) and Poisson INAR(1) models (see also Al-Osh and Alzaid (1987); and, Alzaid and Al-Osh (1988)). Moreover, another approach to addressing the “*multiplication problem*” is the Interger-Valued Generalized AR Conditional Heteroscedasticity (INGARCH) by Ferland et al. (2006), which models the count time series as a linear regression of the conditional means. So that, the count at a given time is generated by using, say, a conditional Poisson distribution (Weiss, 2018). This model led Fokianos et al. (2009) to develop the Poisson AR (PoisAR) model, where the log-linear version of it (Fokianos and Tjøstheim, 2011) is the one used in this paper. Other related works were studied by Fokianos (2012), and Fokianos and Fried (2012).

One thing to emphasize, however, is that this paper further extends the work of Fokianos and Tjøstheim (2011) by also proposing a Bayesian approach to it, which to the best knowledge of the

author none has ever explored yet. The direct benefit of using Bayesian inference is the ability to study the behavior of the parameters through its marginal posterior distribution. This distribution allows researchers to study the uncertainty of the parameters, and give insights into the full characteristics of the estimate. Most importantly, it naturally provide prediction interval due to probabilistic nature of Bayesian inference, and is very useful for estimating the TC activities. Further, the work of Brandt and Sandler (2012) on Bayesian Poisson Vector Autoregressive (Bayesian PoisVAR) model is based on a different specification, and is not the same model (at least the univariate counterpart of PoisVAR) with the one by Fokianos and Tjøstheim (2011), which is extended to Bayesian in this paper.

Moving on, in terms of the data, the interest of the paper is not on the broader region of WNP, but rather on a specific country, the Philippines, in particular, the Philippine Area of Responsibility (PAR). Hence, the frequency defined here is not based on the TC geneses—since these activities often developed outside PAR—but rather on the actual TCs entering the PAR. This study can therefore be compared to one of the locations studied by Magee et al. (2021), the Philippines.

As for the covariates, Magee et al. (2021) summarized the indices often used by their automated variable selection algorithm as predictors for the Philippine TC counts. This is the first study to have come up with indices for the Philippines, as far as we know. The indices identified are: NINO3.4 (average SSTs in the NINO3.4 region defined over  $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ,  $170^{\circ}\text{W}$ – $120^{\circ}\text{W}$ ), IOD E (SST anomalies in the IOD E region defined over  $0^{\circ}\text{S}$ – $10^{\circ}\text{S}$ ,  $90^{\circ}\text{E}$ – $110^{\circ}\text{E}$ ), IOD W (SST anomalies in the IOD W region defined over western pole of the DMI at  $10^{\circ}\text{N}$ – $10^{\circ}\text{S}$ ,  $50^{\circ}\text{E}$ – $70^{\circ}\text{E}$ ), DMI (difference in SST anomalies between IOD W and IOD E), PNA (Rotated Principal Component Analysis (RPCA) applied to monthly standardised 500mb height anomalies between  $20^{\circ}\text{N}$  and  $90^{\circ}\text{N}$ ), QBO (zonal average of the equatorial 30mb zonal wind) and PMM (Maximum Covariance Analysis (MCA) to SSTs at  $21^{\circ}\text{S}$ – $32^{\circ}\text{N}$ ,  $74^{\circ}\text{W}$ – $15^{\circ}\text{E}$ ). These covariates, however, are not explored in this paper. Instead, this work aims to provide a simpler model that is easy to operationalize but with competitive performance. By this, we mean, using historical data as the only predictors, autoregressed against its future values via a canonical link function, in this case the log function.

The absence of independent covariates makes the proposed model simpler if not the simplest compared to most models proposed in the literatures mentioned earlier. It comes with an obvious cost, the model error. This is especially a problem for annual resolution since most temporal patterns are loss due to aggregation. In order to address this, the study proposes modeling at a semester-level granularity instead. This follows from the study of Asaad (2021) for the case of the Philippines, where the first semester (January to June) is not a TC season, so zero to very few typhoons are recorded; whereas, the second semester is where most of the typhoons occur. Therefore, the first semester is expected to have lower TC totals, while the second semester is expected to have higher frequencies. This annually repeating pattern is suitable for autoregressive (AR) models since it uses past values as its indicator for forecasting future values. Further, most literatures mentioned, only model typhoon season months, which in the case of the PAR corresponds to the second semester. However, the benefit of including the first semester which are the non-TC seasons follows from the assumption that, the first semester's variations can affect the second semester's variations; and, that the second semester's variations can affect the succeeding first semester's variations. This assumption is based on the dynamics of the weather discussed earlier, where current state of the weather can be explained by its preceding state. Moreover, the simplicity of the model allows it to be easy to use in terms of operationalization. Lastly, Magee et al. (2021) used June–November as TC season for the Philippines, which more or less the second semester considered here.

As for the Bayesian inference proposed here, Chand et al. (2010) also used Bayesian approach for Poisson regression modeling of the TC activities in the Fiji region. The authors used Gibbs sampling as the choice for Markov Chain Monte Carlo (MCMC) (Hastings, 1970) algorithm for 10000

iterations, with 3000 burn-in. They also compared the result with the Maximum Likelihood Estimate (MLE) of the Frequentist approach. Further, Werner and Holbrook (2011) used Bayesian model for the TC formation in the Australian region using Gaussian distribution as *a prioris* for the parameters, and used Slice sampling (Neal, 2003) as the MCMC algorithm. We emphasize that, while the said papers mentioned used Bayesian inference for Poisson regression, they did not used Bayesian Poisson *autoregression* model proposed in this paper.

In terms of the MCMC algorithm, new innovations have been made in the past decade. Among these is the incorporation of physical dynamics to the sampler to help assists in generating proposal steps when exploring the posterior distribution, for example the Hamiltonian Monte Carlo (HMC) (Neal, 2011) algorithm uses Hamiltonian dynamics to properly explore or sample the target distribution. With this success, HMC have been extended to address other limitations as well, for example the Stochastic Gradient HMC by Chen et al. (2014) aimed at applying HMC to large datasets, which Asaad and Magadia (2019) applied to Bayesian time series modeling. For this paper, the data is not that large enough to at least notice computationally. Hence, another recent extension to HMC is used instead, the No-U-Turn Sampler (NUTS) sampling algorithm proposed by Hoffman and Gelman (2014), which extends HMC to eliminate the need to set the hyperparameter  $L$  (number of Leapfrog steps).

In summary, this paper has three main contributions to the literature of TC activities: *i.* the use of PoisAR model based on the specification of Fokianos and Tjøstheim (2011); *ii.* the extension of PoisAR to Bayesian framework; and, *iii.* the first TC count model for the Philippines with prediction intervals. Lastly, since the study proposes a new methodology for forecasting TC frequencies, details are given for TC researchers and those dealing with count time series modeling in other domains.

## 2 Data and methods

The data is publicly available from the Japan Meteorological Agency's (JMA's) Regional Specialized Meteorological Center (RSMC). The available time points run from 1951 to first semester of 2021. In total, there are 1852 unique TC international IDs (IDs).

In order to count the frequency of the TCs for the PAR, the 1852 TC IDs are further filtered to only those that have entered the PAR. These include tracks that made a recursive path, meaning entering the PAR briefly and left for the north. In total, 1250 TC IDs tallied under the PAR. It should be noted that, the Philippines has its own local weather agency (PAGASA short for Philippine Atmospheric, Geophysical and Astronomical Services Administration) tracking the TCs, and there can be minor differences in terms of TC counts with that in JMA. Nonetheless, the difference should be marginal. Fig. 1a shows the 1852 TCs in the WNP, and Fig. 1b shows the 1250 TCs in PAR (tails and heads of the tracks beyond PAR were cut out for emphasis). These data points were then grouped per semester annually, so that Fig. 2 depicts the time series of the historical frequencies of the TCs in PAR. It should be noted that Asaad (2021) defines the Philippine storms as those that went straight to the Philippines, and classified those that made a recursive track as belonging to a different cluster, including those that briefly entered PAR. While it differ, the proposed definition in this paper is already a good proxy for the number of storms that passed straight through the country.

In this study, the data is partitioned into training and testing, with training accounting for 80% of the data, that is up until the first semester of 2007, and the rest accounts for the test data (up until the first semester of 2021). The model parameters are estimated using the training data, and then validated in the test data. The software used for this study is Julia (Bezanson et al., 2017) programming language, using Turing.jl (Ge et al., 2018) as the core library for Bayesian inference. MCMC diagnostics using Gelman-Rubin statistics is done using R (R Core Team, 2020) programming language but is executed within Julia using RCall.jl.

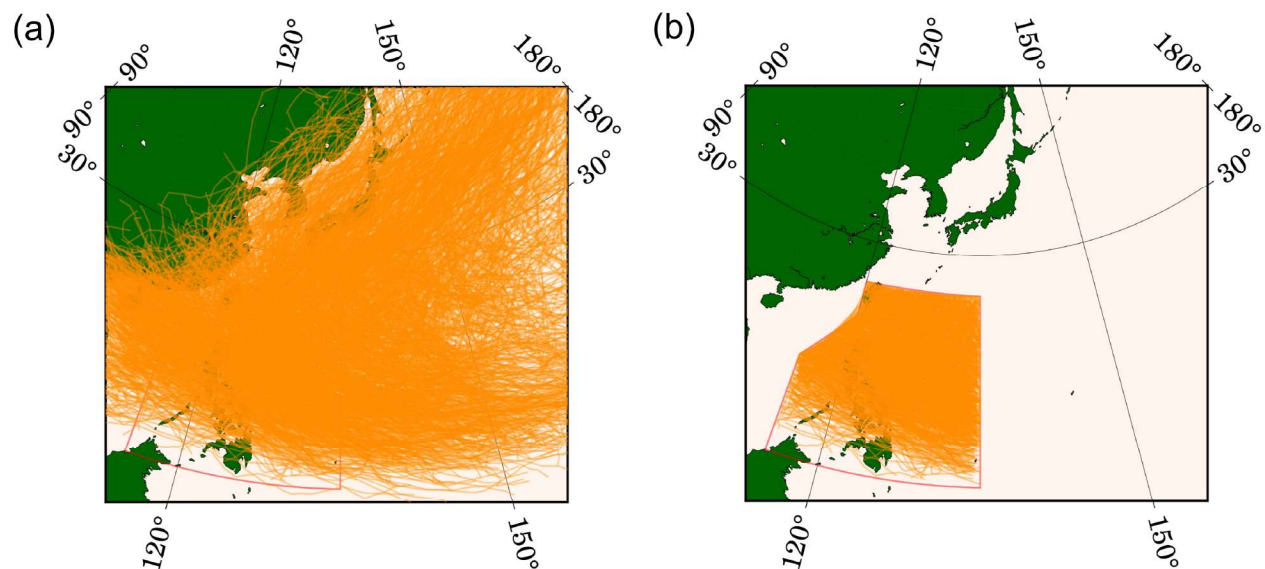


Fig. 1: Typhoon tracks in the **a** Western North Pacific (WNP) region; and, **b** the Philippine Area of Responsibility (PAR)

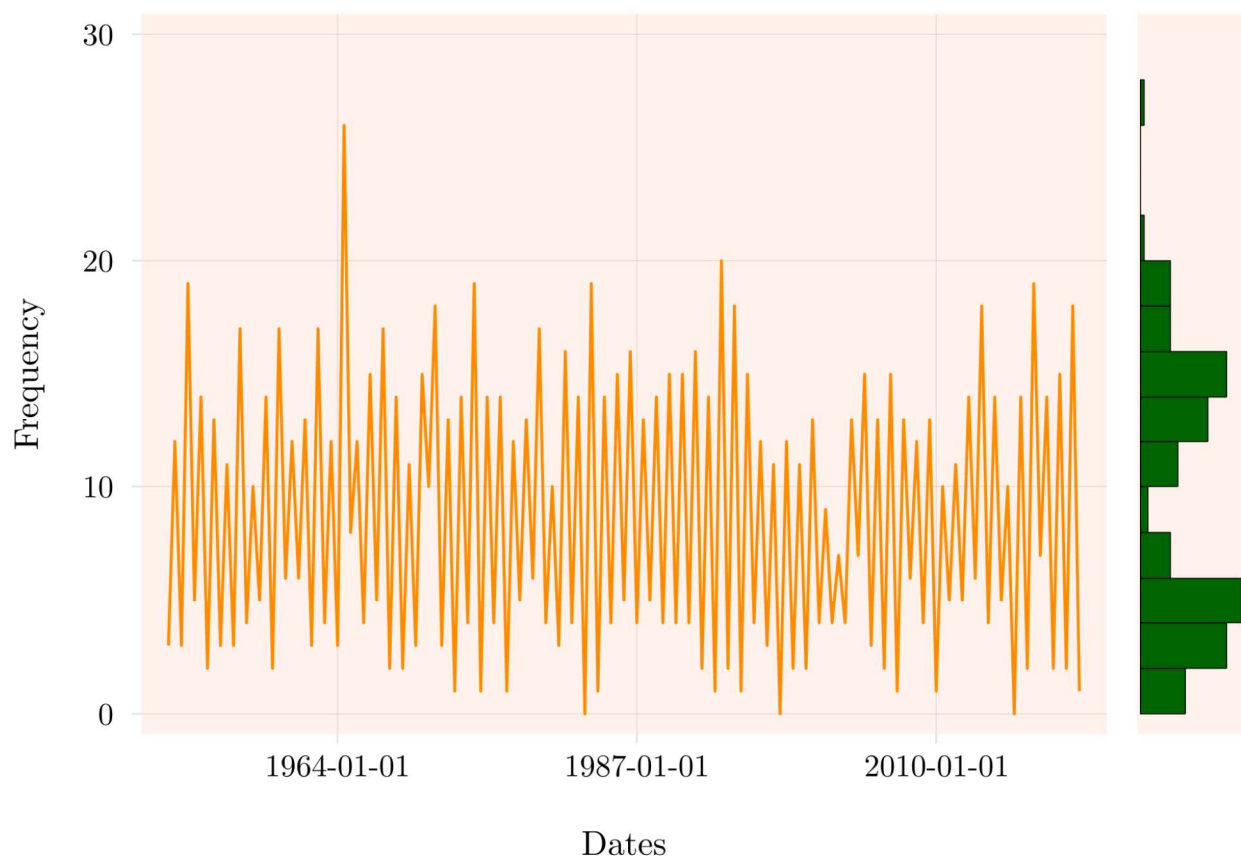


Fig. 2: Tropical cyclone historical frequencies in semesters recorded from 1951 to 2021 under the Philippine Area of Responsibility (PAR)

The remainder of the paper is organized as follows: Section 2.1 introduces the log-linear PoisAR model; Section 2.2 specifies the prior distribution to the parameters of the model; Section 2.3 discusses the model estimation using Bayesian inference; Section 2.4 tackles the necessary diagnostics for the Markov chains; Section 2.5 discusses the model inference; Section 3 presents the results; and, Section 4 is the conclusion and future works.

## 2.1 Log-linear PoisAR model

The log-linear PoisAR model proposed by Fokianos and Tjøstheim (2011) is defined in Definition 1.

**Definition 1 (Log-linear PoisAR model).** Let  $\mathcal{F}_t^{Y,\eta} := \sigma(Y_s, s \leq t, \eta_0)$  be the  $\sigma$ -algebra generated by  $\{Y_0, Y_1, \dots, Y_t, \eta_0\}$ , such that  $Y_t | \mathcal{F}_{t-1}^{Y,\eta} \sim \text{Poisson}(\lambda_t)$ , then the log-linear PoisAR model is defined by

$$\eta_t = \kappa + \sum_{m=1}^b \alpha_m \eta_{t-m} + \sum_{l=1}^p \varphi_l \log(Y_{t-l} + 1), \quad (1)$$

where  $\eta_t := \log \lambda_t$ ;  $\kappa, \alpha_m, \varphi_l \in \mathbb{R}$ ; and,  $t \in \mathbb{N}$ .

*Remark 1.* The  $\eta_{t-q}$  in Equation (1) is the feedback mechanism of the model.

The choice for the appropriate temporal lag of  $Y_t$  can be determined using the sample Autocorrelation Function (ACF). The paper, however, settles on temporal lag 2, mainly because the current variations of the frequencies are assumed to be well explained by the preceding two semesters, which form a contrasting (first semester TC totals is always lower than the second semester) effect, annually. Further, the historical data at the semester-level is periodic as shown in Fig. 2, making the ACF decays slowly due to high correlations between temporal lags. As for the feedback mechanism, the paper aims for parsimony and sets it to lag 1. Hence, the model considered in this paper is given by

$$\eta_t = \kappa + \alpha \eta_{t-1} + \varphi_1 \log(Y_{t-1} + 1) + \varphi_2 \log(Y_{t-2} + 1). \quad (2)$$

The addition of 1 inside the  $\log(Y_{t-i} + 1)$ ,  $i \in \{1, 2\}$ , is meant to handle 0 TC totals as Fokianos and Tjøstheim (2011) pointed out. This is particularly true for the Philippines, where the first semester (non-TC season) can have 0 TC totals, as seen in Fig. 2.

## 2.2 Prior distribution specification

The model in Equation (2) is estimated using Bayesian inference. This is done through the Bayes' Theorem, which requires prior distribution specification on the target parameters. For this paper, the following are the *a priors*:

$$\kappa \sim \text{Uniform}(0, 15) \quad (3)$$

$$\alpha | \sigma_1 \sim \text{Normal}(0, \sigma_1) \quad (4)$$

$$\varphi_1 | \sigma_2 \sim \text{Normal}(-1, \sigma_2) \quad (5)$$

$$\varphi_2 | \sigma_3 \sim \text{Normal}(1, \sigma_3) \quad (6)$$

$$\sigma_1 \sim \text{InverseGamma}(3, 1) \quad (7)$$

$$\sigma_2 \sim \text{InverseGamma}(3, 1) \quad (8)$$

$$\sigma_3 \sim \text{InverseGamma}(3, 1). \quad (9)$$

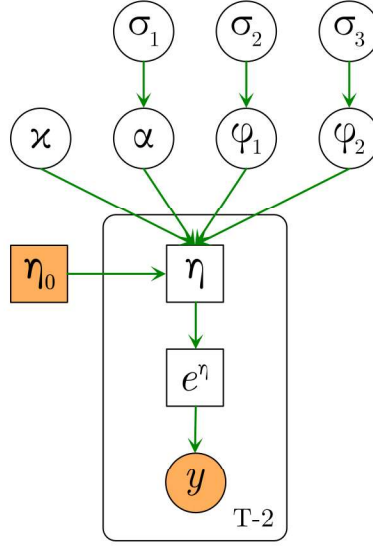


Fig. 3: Plate diagram of the proposed hierarchical Bayesian model

The specification above assumes that  $\varphi_1$  and  $\varphi_2$  have opposite effects since the preceding two frequencies as predictors are expected to have contrasting values, hence the difference on the mean of their *a priori*s. The  $\kappa$ , which serves as the constant of the model, is hoped to compensate the needed values to get close to the target frequencies. This is therefore assigned to a Uniform distribution over the domain 0 to 15. On the other hand,  $\alpha$  is simply assumed to be Gaussian distributed with location parameter at 0. Finally, the hyperparameters  $\sigma_1, \sigma_2$  and  $\sigma_3$  are simply initialized to Inverse Gamma with parameters 3 and 1. So that, Fig. 3 is the corresponding plate diagram of the proposed hierarchical Bayesian model. From the said figure, it should be understood that there are  $T - 2$  plates since the log-linear model proposed is of order 2 (see Equation (2)). Further, the square shapes in Fig. 3 are constants and circles are probability distributions. Shaded shapes implies that they are known, whereas unshaded ones are unknown. Those that fall within the big rectangular (the *plate*) shape implies that these variables vary across time.

### 2.3 Model estimation

As mentioned above, inference on the target parameters is done via Bayes' Theorem, which state that for any random variables  $X$  and  $Y$ ,

$$\Pr(Y | X) = \frac{\Pr(X | Y)\Pr(Y)}{\Pr(X)}. \quad (10)$$

Applying this formula to the problem, let  $\theta_1 := [\kappa, \alpha, \varphi_1, \varphi_2]^T$  be the vector of parameters and  $\theta_2 := [\sigma_1, \sigma_2, \sigma_3]^T$  be the vector of hyperparameters. Further, let  $Y_{\leq T}$  be the data such that  $Y_t | \mathcal{F}_{t-1}^{Y, \eta} \sim \text{Poisson}(\exp(\eta_t))$ , so that  $\mathbf{y} = [y_1, \dots, y_t]^T$ , then the posterior distribution is

$$\Pr(\theta_1, \theta_2 | \mathbf{y}) = \frac{\Pr(\mathbf{y} | \theta_1, \theta_2)\Pr(\theta_1, \theta_2)}{\Pr(\mathbf{y})} \quad (11)$$

$$= \frac{\Pr(\mathbf{y} | \theta_1)\Pr(\theta_1 | \theta_2)\Pr(\theta_2)}{\Pr(\mathbf{y})} \quad (12)$$

$$= \frac{\Pr(\mathbf{y} | \theta_1)\Pr(\theta_1 | \theta_2)\Pr(\theta_2)}{\int_{\forall \theta_2} \int_{\forall \theta_1} \Pr(\mathbf{y} | \theta_1)\Pr(\theta_1 | \theta_2)\Pr(\theta_2) d\theta_1 d\theta_2} \quad (13)$$

where

$$\Pr(\mathbf{y} \mid \boldsymbol{\theta}_1) = \prod_{t=3}^T \frac{\lambda_t^{y_t} \exp(-\lambda_t)}{y_t!}, \quad (14)$$

and  $\lambda_t = \exp(\kappa) \lambda_{t-1}^\alpha (y_{t-1} + 1)^{\varphi_1} (y_{t-2} + 1)^{\varphi_2}$  since  $\eta_t := \log(\lambda_t)$  from Definition 1. Further,

$$\Pr(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2) = \Pr(\alpha \mid \sigma_1) \Pr(\varphi_1 \mid \sigma_2) \Pr(\varphi_2 \mid \sigma_3) \Pr(\kappa) \quad (15)$$

$$\Pr(\boldsymbol{\theta}_2) = \Pr(\sigma_1) \Pr(\sigma_2) \Pr(\sigma_3). \quad (16)$$

Finally, the following are the corresponding prior distributions specified in Section 2.2:

$$\Pr(\alpha \mid \mu = 0, \sigma_1) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left(-\frac{\alpha^2}{2\sigma_1^2}\right) \quad (17)$$

$$\Pr(\varphi_1 \mid \mu = 0, \sigma_2) = \frac{1}{\sigma_2 \sqrt{2\pi}} \exp\left(-\frac{\varphi_1^2}{2\sigma_2^2}\right) \quad (18)$$

$$\Pr(\varphi_2 \mid \mu = 0, \sigma_3) = \frac{1}{\sigma_3 \sqrt{2\pi}} \exp\left(-\frac{\varphi_2^2}{2\sigma_3^2}\right) \quad (19)$$

$$\Pr(\sigma_i \mid a = 3, b = 1) = \frac{1}{\Gamma(3)} \sigma_i^{-4} \exp\left(-\frac{1}{\sigma_i}\right), \quad i \in \{1, 2, 3\} \quad (20)$$

$$\Pr(\kappa \mid a = 0, b = 15) = \begin{cases} \frac{k}{15}, & a \leq k \leq b \\ 0, & \text{otherwise} \end{cases}. \quad (21)$$

The main problem with Equation (11) is the high-dimensional integration needed in the denominator, which often is intractable or simply difficult to simplify. Hence, it is often the practice to just approximate  $\Pr(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \mid y_t)$  using MCMC (Hastings, 1970). There are several MCMC algorithms, but the one used here is the NUTS (Hoffman and Gelman, 2014) which is an extension of HMC (Neal, 2011).

The NUTS algorithm requires the gradient of the log of the numerator with respect to the parameters, that is,

$$\frac{\partial}{\partial \theta_i} \log [\Pr(y_t \mid \boldsymbol{\theta}_1) \Pr(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2) \Pr(\boldsymbol{\theta}_2)], \quad i \in \{1, 2\}. \quad (22)$$

These differentiations are handled using the *automatic differentiation* (Griewank and Walther, 2008) built-in within Turing.jl (Ge et al., 2018), which fed the resulting derivative into its NUTS algorithm as defined in Algorithm 6 of Hoffman and Gelman (2014) at p.1611.

In practice, the MCMC algorithms needs few iterations to “warm up,” (also known as *burn-in* portion) and hence a portion of the Markov chain is discarded, leaving with what assumed to be the sample points from the marginal posterior distribution. Moreover, it is recommended to run several Markov chains in parallel to make sure that the algorithm converges to a stationary distribution. For this study, 4 Markov chains are sampled each having 20000 sample size. The burn-in is 10000, leaving 10000 samples for each Markov chains used as the data points for estimating the marginal posterior distributions.

## 2.4 Markov chain diagnostics

MCMC algorithms work in principle by sampling from the *a posteriori* through random walk. The realized independent Markov chains from parallel random walks (with mechanism for accepting/rejecting the samples) are assumed to have converged to some stationary distribution—with burn-in samples

already removed. The popular statistical test for checking the convergence of multiple Markov chains is the Gelman-Rubin statistics proposed by Gelman and Rubin (1992) and further refined by Brooks and Gelman (1998). It should be noted that, Chand et al. (2010) diagnosed the model convergence through repeated samplings with different values of initial conditions, and then checking if the KDEs are identical (meaning the chains converged). As for Werner and Holbrook (2011), they argued that the relatively short (500 iterations) iterative burn-in they used achieves a quick convergence. Both literatures, however, did not use Gelman-Rubin statistics.

Gelman-Rubin statistics works by comparing the estimated between-variance and within-variance of the Markov chains. In particular, the between-variance is the variance of the averages of the independent Markov chains, while the within-variance is the average of the variances of the independent Markov chains. Hence, if the Markov chains did converged, then the two statistics mentioned must approximately equate to each other. For this study, the parameters  $\kappa, \alpha, \varphi_1$  and  $\varphi_2$ ; and, the hyper-parameters  $\sigma_1, \sigma_2$  and  $\sigma_3$  are each sampled from the posterior distribution using 4 parallel Markov chains. These independent chains are then tested for convergence using the Gelman-Rubin's statistics.

## 2.5 Model inference

Given the sampling distributions of the marginal *a posteriori* presented in Section 3.1, inference is done as follows: *i.* for each target parameter, take the average of its 4 independent Markov chains, leaving only one averaged Markov chain with 20000 samples; *ii.* remove the first 10000 samples in the Markov chain as part of the burn-in, leaving 10000 samples; *iii.* use these samples to extract the characteristics of the corresponding marginal *a posteriori*. So that mathematically, let  $N = 10000$ ,  $\kappa := [\kappa_1, \dots, \kappa_N]^\top$ ,  $\alpha := [\alpha_1, \dots, \alpha_N]^\top$ ,  $\varphi_1 := [\varphi_{1,1}, \dots, \varphi_{1,N}]^\top$ , and  $\varphi_2 := [\varphi_{2,1}, \dots, \varphi_{2,N}]^\top$ , then the model is

$$\eta_t := \kappa + \alpha \odot \eta_{t-1} + \varphi_1 \log(y_{t-1} + 1) + \varphi_2 \log(y_{t-2} + 1) \quad (23)$$

$$\hat{y}_t := \lambda_t \quad (24)$$

where  $\eta_t = [\eta_{1t}, \dots, \eta_{Nt}]^\top$ ,  $\lambda_t = [\exp(\eta_{1t}), \dots, \exp(\eta_{Nt})]^\top$  and  $\alpha \odot \eta_{t-1}$  is the one-to-one element-wise multiplication between  $\alpha$  and  $\eta_{t-1}$ . The resulting inference from Equation (24) is  $\hat{y}_t$ , a vector of estimated frequencies at time  $t$ . The final point forecast is therefore obtained by simply averaging the vector, that is,  $\hat{y}_t := \sum_i \exp(\eta_{it})$ . The vector  $\hat{y}_t$  describes the uncertainty of the mean forecast  $\hat{y}_t$ . This uncertainty is referred in here as the *epistemic uncertainty* (first coined by Kendall and Gal (2017) to the best knowledge of the author). This is because  $\hat{y}_t$  is a realization of our hypothesis or belief about the function that generated the data. The specified hypothesis assumes that the data was generated by a log-linear model (i.e., Equation (2)). Hence, the variance of the elements of the vector  $\hat{y}_t$  describes the uncertainty of this belief.

One thing to note, however, is that Equation (23) assumes that  $y_{t-1}$  and  $y_{t-2}$  as known values. That is, the inferred  $\hat{y}_t$  is not used as input to forecast  $\hat{y}_{t+1}$ . For such cases, the type of forecast is referred in here as *rolling forecast*, which will be discussed later in this section. Moreover,  $\eta_{t-1}$  is a vector for  $t \geq 2$ , since for  $t = 1$ ,  $\eta_{t-1} = \eta_0$  which is a scalar. In this case, the  $\eta_t$  is used to infer  $\eta_{t+1}$ , making  $\eta_t$  as the feedback mechanism for  $t + 1$ .

Moreover,  $\hat{y}_t$  is the forecasts conditional means of the observation  $y_t$ , and because  $y_t$  is assumed to be Poisson distributed, it therefore meant that  $\hat{y}_t | \mathcal{F}_{t-1}^{Y, \eta} \sim \text{Poisson}(\hat{y}_{it}), \forall i, t$ . The variance of this Poisson distribution describes the uncertainty of the predicted observations, which Kendall and Gal (2017) named as *aleatoric uncertainty*, and is also adopted here.

Since for every element of  $\hat{y}_t$  corresponds to a predicted mean, then there will be multiple quantiles/percentiles of the Poisson distribution at time  $t$ , each having mean corresponding to the elements

of  $\hat{y}_t$ . To properly display this, we first compute for the following order statistics: the maximum  $\hat{y}_{(N)t}$  and the minimum  $\hat{y}_{(0)t}$  for all  $t$ . That is,

$$\hat{y}_{(N)t} := \max\{\hat{y}_{1t}, \dots, \hat{y}_{Nt}\} \quad (25)$$

$$\hat{y}_{(0)t} := \min\{\hat{y}_{1t}, \dots, \hat{y}_{Nt}\}. \quad (26)$$

We then compute the prediction intervals for 90% (lower bound at 2.5th percentile and upper bound at 97.5th percentile), 75% (12.5th lower and 87.5th upper percentiles), and 50% (25th lower and 75th upper percentiles). The percentile/quantile  $q$  is computed as follows:  $\hat{y}_t^{(q)} := F^{-1}(q \mid \hat{y}_{(0)t})$  for percentiles below the median (i.e., 50th percentile); and,  $\hat{y}_t^{(q)} := F^{-1}(q \mid \hat{y}_{(N)t})$  for percentiles above the median. The  $F^{-1}(q \mid \lambda)$  is the image of  $F(\hat{y}_t^{(q)} \mid \lambda)$  defined below:

$$F(\hat{y}_t^{(q)} \mid \lambda) := \Pr(Y \leq \hat{y}_t^{(q)} \mid \lambda) = \sum_{i=1}^{\hat{y}_t^{(q)}} \frac{\lambda^i \exp(-\lambda)}{i!}. \quad (27)$$

Moving on, the main challenge of doing rolling forecast is due to the cost it entails on models with independent covariates, which in literatures reviewed here are often given as indices. It is costly in the sense that it requires the model to have the covariates be available as well on the future time points to be forecast, which means these predictors need to be modeled as well. This is not a problem for models with no independent covariates, however, since it only depends on its past values like the one used here. To do this, the following iterative forecasting is used:

$$\eta_t := \kappa + \alpha \odot \eta_{t-1} + \varphi_1 \odot \mathbf{z}_{t-1} + \varphi_2 \odot \mathbf{z}_{t-2} \quad (28)$$

$$\hat{y}_t := \lambda_t, \quad (29)$$

where  $\mathbf{z}_{t-l} := [\log(\hat{y}_{i(t-l)} + 1), \dots, \log(\hat{y}_{N(t-l)} + 1)]^\top, \forall l \in \{1, 2\}$ . The main difference between Equation (28) and Equation (23) is that the former uses the forecast  $\hat{y}_t$  as input to forecast  $\hat{y}_{t+1}$ , which means  $y_t$  is not available to forecast  $t + 1$ . Rolling forecasting should be taken with caution, however, since it can quickly accumulate the model residuals with respect to the length of the target horizon. Hence, the proposed model in this paper is only recommended for one-semester ahead forecasting using Equation (23).

Finally,  $\eta_0$  evolved from forecasting the training dataset, can also be used as the initial feedback mechanism for the test data. This is contrary to simply fixing it to 1, as was used in estimating the posterior. Nonetheless, both specification will give almost the same results.

### 3 Results and discussions

The results are presented into the following sections: Section 3.1 discusses the kernel density estimates (KDEs) of the sampling distributions of the marginal *a posteriori* of the parameters; Section 3.2 presents the statistical tests on the convergence of the Markov chains; Section 3.3 discusses the sensitivity of the model with respect to  $\eta_0$ ; and, Section 3.4 discusses the corresponding forecast of the model.

#### 3.1 Marginal posterior distributions

The statistics of the KDEs are given in Table 1 with corresponding plots provided in Fig. 4. More specifically, Fig. 4a-g corresponds to the KDEs of the parameters  $\kappa, \alpha, \varphi_1, \varphi_2, \sigma_1, \sigma_2$  and  $\sigma_3$ , respectively. Starting with the KDEs of the  $\kappa$ , the marginal distribution seem to follow a Gaussian density

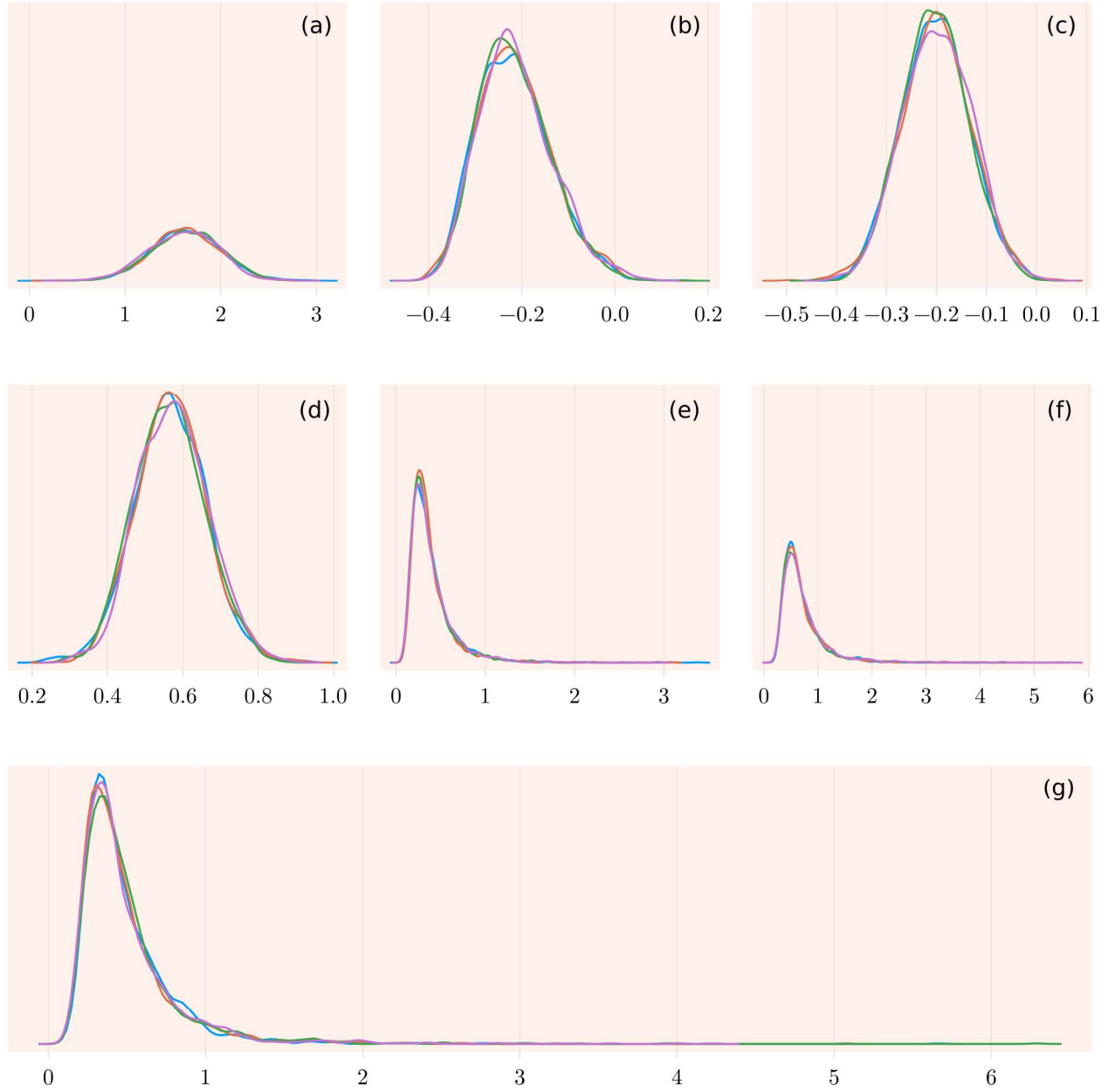


Fig. 4: Kernel density estimates of the *a posteriori* of the parameters: **a**  $\kappa$ ; **b**  $\alpha$ ; **c**  $\varphi_1$ ; **d**  $\varphi_2$ ; **e**  $\sigma_1$ ; **f**  $\sigma_2$ ; and, **g**  $\sigma_3$

function centered around 1.57, with estimated standard deviation of 0.21 (refer to standard error (SE) in Table 1). Interestingly, the NUTS algorithm managed to explore the marginal of  $\kappa$  despite the initial prior distribution assumed was  $\text{Uniform}(0, 15)$ , that is, the range should go all the way up to 15, but is now significantly shrank after observing all observations. Further, the parameter  $\alpha$  which was initially set to a Gaussian *a priori* centered at 0, now slightly moves to the negative values with an estimated mean of -0.22 and a standard error of 0.06.

As for the KDEs of  $\varphi_1$  and  $\varphi_2$ , both have contrasting supports, that is,  $\varphi_1$  has negative domain centered at -0.19, while  $\varphi_2$  is on the positive side with mean 0.59. One can also observed that, both  $\varphi_1$  and  $\varphi_2$  like  $\kappa$  and  $\alpha$  have symmetric curve with skewness near 0. Finally, the three hyperparameters,  $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$  have marginal *a posteriori* closely resembling a Inverse Gamma distribution.

Param	Percentiles					Mean	SE	Skew.	Kurt.
	5th	25th	50th	75th	95th				
$\kappa$	1.33	1.51	1.64	1.75	1.94	1.63	0.18	0.03	0.0
$\alpha$	-0.28	-0.24	-0.22	-0.19	-0.15	-0.21	0.04	0.27	-0.15
$\varphi_1$	-0.26	-0.22	-0.2	-0.18	-0.14	-0.2	0.03	0.01	-0.02
$\varphi_2$	0.49	0.54	0.57	0.6	0.65	0.57	0.05	0.02	0.04
$\sigma_1$	0.24	0.3	0.36	0.45	0.62	0.39	0.13	1.7	5.99
$\sigma_2$	0.46	0.56	0.66	0.79	1.07	0.7	0.2	1.77	5.54
$\sigma_3$	0.31	0.39	0.46	0.56	0.79	0.5	0.16	1.98	7.65

Table 1: Statistics of the kernel density estimates of the marginal sampling distributions

### 3.2 Markov chain diagnostics

The Gelman-Rubin statistics (Gelman and Rubin, 1992; Brooks and Gelman, 1998) returns a  $\mathbb{R}$ -valued shrink factor. The shrink factor indicates that, if the test statistics of the Markov chains of a given parameter is 1, then the between-variance and within-variance (see discussions in Section 2.4) are equal, meaning the parallel Markov chains tested have all converged to a stationary distribution. For the proposed model, the Markov chains of the estimates of the parameters have Gelman-Rubin statistics equal to 1.

To see this visually, Fig. 5 shows the evolution of the shrink factor across iterations. Towards the end, both the median and even the 97.5% percentile of the shrink factor are already 1 or near 1. The convergence is usually determined if the shrink factor is less than 1.1 (the “*stringent standard*,” see Brooks and Gelman (1998)). Fig. 5 helps users to specify the ideal burn-in. This study, however, settled on 10000 samples as part of the burn-in, leaving 10000 remaining samples as estimate for the marginal distribution. Note that, these remaining 10000 samples account for 1 Markov chain only. Therefore, for the 4 Markov chains ran for each parameter, the total samples used to estimate the marginal distribution is  $10000 \times 4 = 40000$ .

### 3.3 Model sensitivity on $\eta_0$

The proposed model requires initialization of its feedback mechanism,  $\eta_t$ . It is therefore important to understand the sensitivity of the model with respect to  $\eta_0$ . To analyze this sensitivity, Fig. 6 shows the variabilities of the model metrics with respect to the values of  $\eta_0 \in \{1, 1.5, 2, 2.5, \dots, 15\}$  on both training and testing data. Fig. 6a-d correspond to the MSE (Mean Squared Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) and Correlation, respectively. One thing to emphasize, are the high errors/low correlations at  $\eta_0 = 2.5$  and  $\eta_0 = 3.0$ . Upon inspection, the model does perform poor under these initializations. However, the corresponding marginal distributions estimated using MCMC are not stationary, that is, the NUTS algorithm cannot explore properly the marginal *a posterioris* for the said  $\eta_0$ s. This is due to the seed fixed for NUTS for all  $\eta_0 \in \{1, 1.5, 2, 2.5, \dots, 15\}$  for reproducibility purposes. So one can argue that, under a different seed, the Turing.jl’s NUTS could have positioned the first step of the Markov chain properly within the target posterior. In spite of that, the rest of the values of the four metrics for other  $\eta_0$  values are not that varied. In fact, excluding the values for  $\eta_0 = 2.5$  and  $\eta_0 = 3.0$ , the coefficient of variations (CVs) of the sensitivity across metrics for

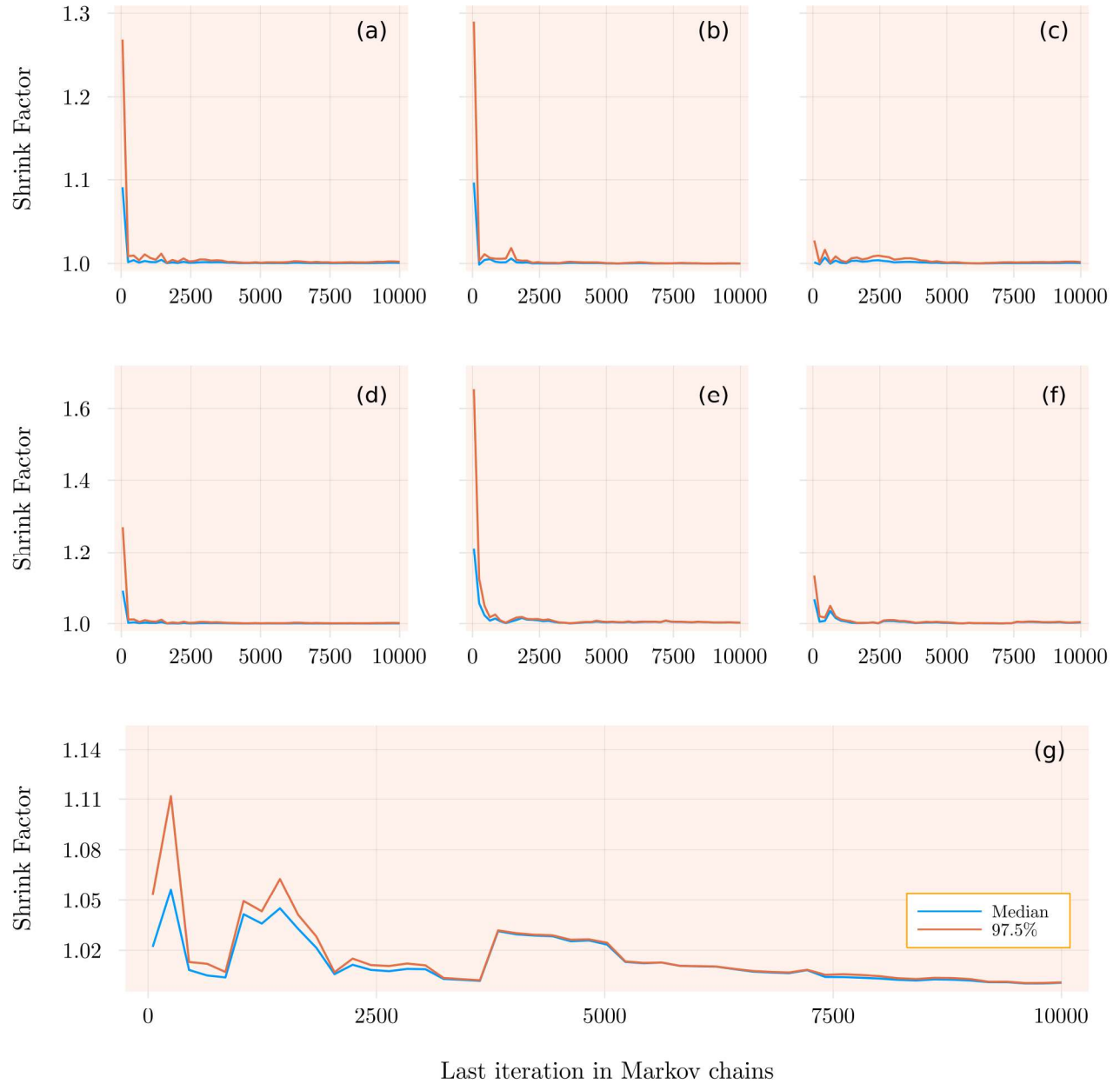


Fig. 5: Evolution of the Gelman-Rubin's shrink factor for test of stationarity for **a**  $\kappa$ ; **b**  $\alpha$ ; **c**  $\varphi_1$ ; **d**  $\varphi_2$ ; **e**  $\sigma_1$ ; **f**  $\sigma_2$ ; and, **g**  $\sigma_3$

the training dataset are as follows: 0.04 (MSE), 0.02 (RMSE), 0.03 (MAE) and 0.01 (Correlations); and, for testing 0.38 (MSE), 0.15 (RMSE), 0.2 (MAE) and 0.13 (Correlations). All of the CVs are small or near 0.

Lastly, the sensitivity under the test data was computed using the post feedback mechanism  $\eta_t$ , evolved from computing the training data, as the initial value of  $\eta_0$  for forecasting the test data. Although not shown here, this is not significantly different from, say, fixing it to  $\eta_0 = 1$  (initially used for estimating the posterior) for validation data. This therefore shows that, in terms of operationalization of the model,  $\eta_0$  can be set to any value, at least within the range studied here.

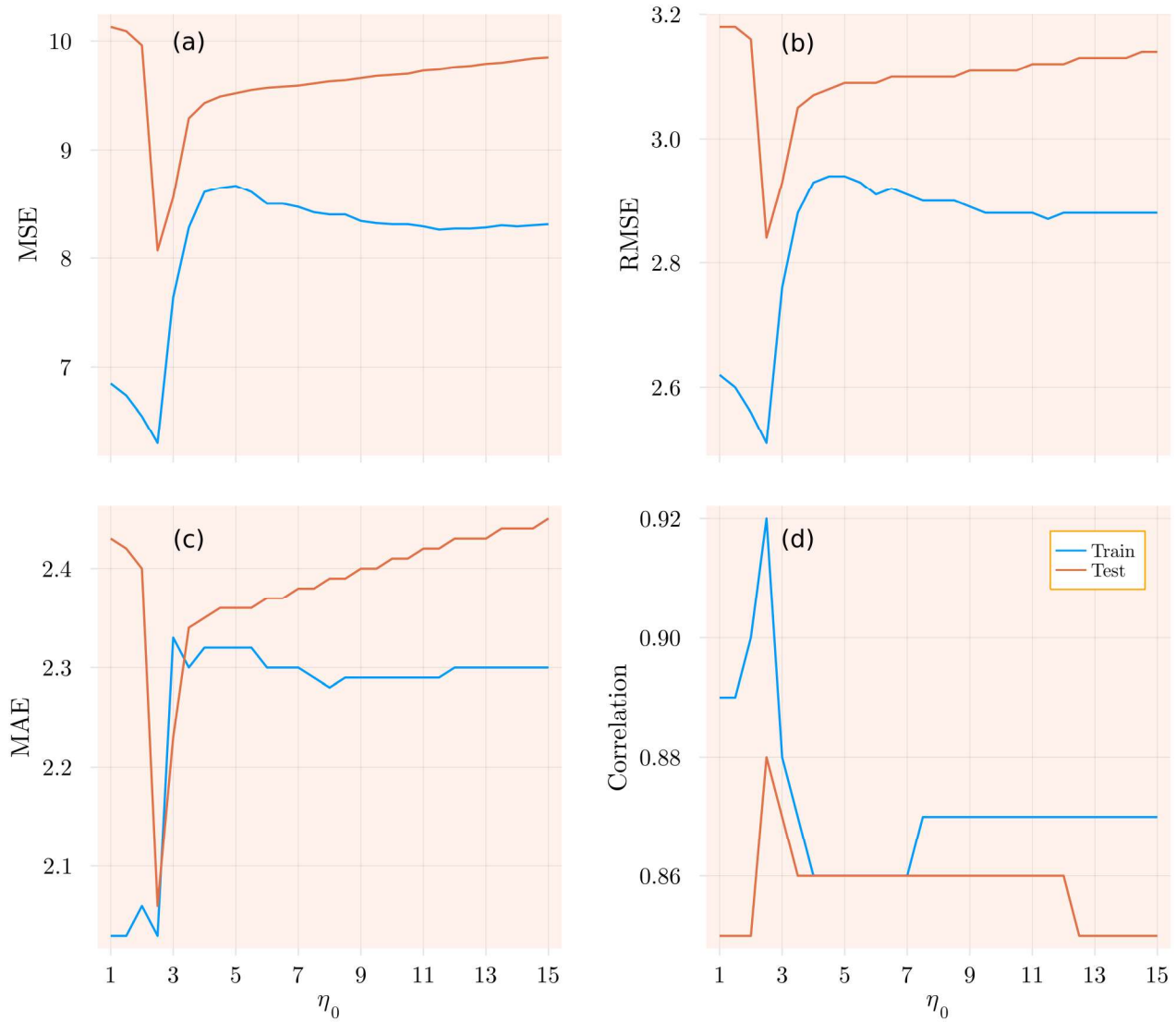


Fig. 6: Sensitivity of the model on the values of  $\eta_0$  on both training and testing data across metrics: **a** MSE; **b** RMSE; **c** MAE; and, **d** Correlation

Data	MSE	RMSE	MAE	Correlation
Train	9.63	3.1	2.39	0.86
Test	8.4	2.9	2.28	0.87

Table 2: Performance of the models on both train and test datasets

### 3.4 Forecasts

The errors of the model are provided in Table 2. The corresponding forecasts are given in Fig. 7 for training data, and Fig. 8 for test data. From both figures, the model does capture well the characteristics of the time series, although it misses the significant spike on the second semester of 1964 in the training dataset. For the testing dataset, it also misses few dips. Despite that, however, the misses are mostly within the prediction intervals as emphasized by the *aleatoric* uncertainty. Further, the first

Year-Sem	Aleatoric Percentiles						Mean	Epistemic	
	2.5	12.5	25	75	87.5	97.5		Lower	Upper
2021 - Sem2	7	10	12	20	22	25	15.64	14.28	17.80
2022 - Sem1	0	1	2	7	8	11	4.45	3.41	5.50
2022 - Sem2	1	2	3	8	10	12	5.05	4.05	6.58

Table 3: Tropical cyclone count forecasts of the proposed model up to the last semesters of 2022

two points shows that the model is not confident with its forecast, this is emphasized by the *epistemic* uncertainty, which is higher compared to other points—with the exception of second semester 2016, where the model also returned a pronounced *epistemic* uncertainty despite capturing the actual frequency in its forecast mean.

Moreover, it is a common practice in the literature (for example Zhang et al. (2018), McDonnell and Holbrook (2004b), among others) of TC count forecasting to also check on the Pearson correlation between the forecast and the actual values. For the proposed model, the correlation is high since the model takes advantage of the semester-level data, where periodic patterns recur every year.

As for the accuracy, while there are differences in terms of the setup of modeling, validation method, to granularity of the data, we'll compare the proposed model to that of Magee et al. (2021) for the Philippines. As for the MAE, their model performed well with error less than 2, which is expected as they used at least 5 independent covariates compared to the proposed model, which uses only the historical data. However, in terms of the correlation, the semester-level data used here benefited the model well, and is higher than that by Magee et al. (2021), which is max at around 0.8. In addition to this, the added prediction intervals by the proposed model can provide more insights that can be useful for decision-making.

In terms of the forecast of first and second semesters of 2022, Table 3 presents the details. The two-ahead forecast horizon is possible by using the mean forecast of the first semester as input to forecast the second semester, that is, a roll forecast was done for the second semester. From the inference, about 4 TCs are expected in the first semester of 2022, although it can be as low as 2 according to the 50% prediction interval. For the second semester, on the other hand, about 5 TCs are expected, but it can be as high as 10 according to the 87.5th percentile—the upper limit of the 75% prediction interval. The low estimate of about 5 TCs for the second semester is obviously an underestimation since the second semester is a TC season, but this is expected when doing a rolling forecast, the forecast will soon converge into a constant value if we roll further to future time points.

Lastly, it can be observed that, for the second semester, the model has greater variability on its *epistemic* uncertainty, with range of 2.53 (6.58 – 4.05) compared to the first semester with 2.09 range. This is expected since the forecast of the second semester is based on a rolling forecast, which accumulates the error from the previous forecast. Nevertheless, the prediction intervals provided by the *aleatoric* percentiles can give more information about the expected TC counts.

## 4 Conclusion and future works

The aim of the study was to explore the use of PoisAR model, in the context of Bayesian inference, for modeling the Philippine storm frequency. The results of the inference suggest that the proposed model, while simple enough compared to those in other TC count literatures, can well describe the characteristics of the Philippine TC frequencies. In particular, the RMSE of the proposed model is

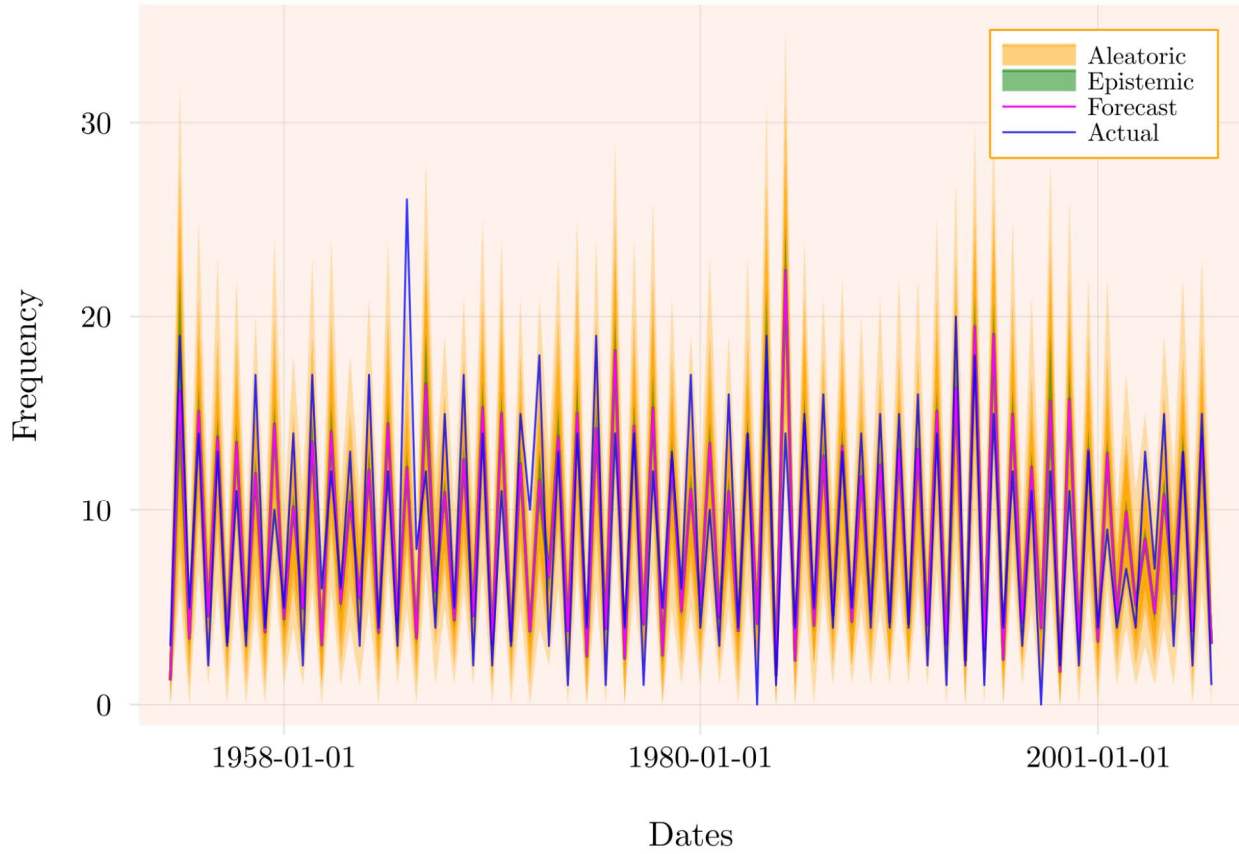


Fig. 7: One-semester ahead forecast on training data.

3.1 for training (80% of the data) and 2.9 for testing (20% of the data). Further, actual frequencies not captured by the forecast mean are mostly within the prediction intervals. In particular, the availability of both *epistemic* and *aleatoric* uncertainties can give insights into the confidence of the model, which are useful for policy-makers. Lastly, the absence of the covariates, meant that the model is easy to maintain and operationalize. We've shown this on forecasting the 2022, where roll forecasting was done to predict the TC count for the second semester. So that, the model inferred about 4 TCs on the first semester of 2022, and about 5 TCs for the second semester.

There are still room for creativity in terms of the proposed model. The first obvious improvement is the use of independent covariates to capture other variations that the model can't explain. Second, the exploration of the non-linear version of PoisAR; and finally, the exploration of multiple horizon forecasting instead of one-point forecast at a time.

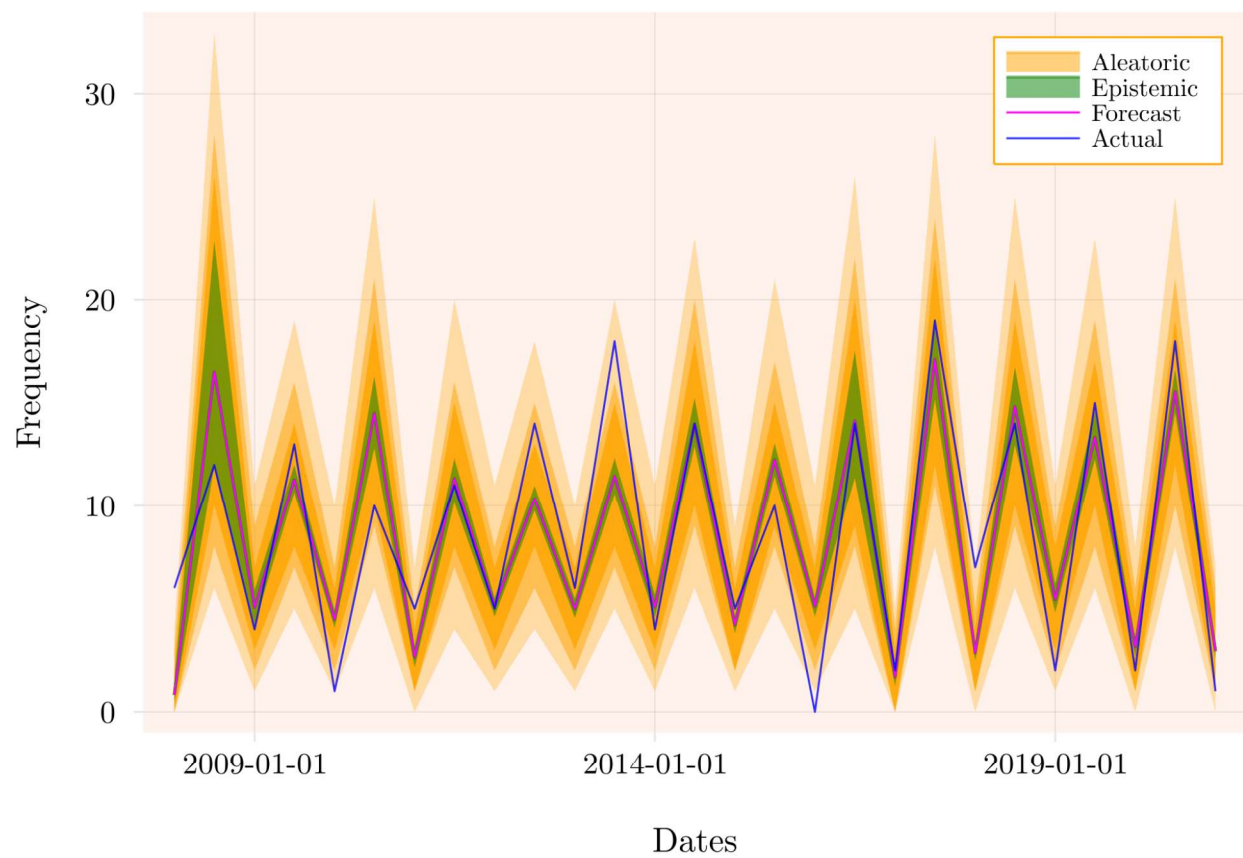


Fig. 8: One-semester ahead forecast on testing data.

## Bibliography

- Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (inar(1)) process. *Journal of Time Series Analysis*, 8(3):261–275.
- Alzaid, A. and Al-Osh, M. (1988). First-order integer-valued autoregressive (inar (1)) process: Distributional and regression properties. *Statistica Neerlandica*, 42(1):53–61.
- Asaad, A.-A. B. (2021). Circular characteristics of the philippine storm tracks. *Spatial Information Research*.
- Asaad, A.-A. B. and Magadia, J. C. (2019). Stochastic gradient hamiltonian monte carlo on bayesian time series modeling. Quezon City, Philippines. 14th National Convention on Statistics.
- Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98.
- Brandt, P. T. and Sandler, T. (2012). A bayesian poisson vector autoregression model. *Political Analysis*, 20(3):292–315.
- Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.
- Chan, J. C. L. (1985). Tropical cyclone activity in the northwest pacific in relation to the el niño/southern oscillation phenomenon. *Monthly Weather Review*, 113(4):599 – 606.
- Chan, J. C. L. (1995). Tropical cyclone activity in the western north pacific in relation to the stratospheric quasi-biennial oscillation. *Monthly Weather Review*, 123(8):2567 – 2571.
- Chan, J. C. L., en Shi, J., and Liu, K. S. (2001). Improvements in the seasonal forecasting of tropical cyclone activity over the western north pacific. *Weather and Forecasting*, 16(4):491 – 498.
- Chan, J. C. L., en Shi, J., and man Lam, C. (1998). Seasonal forecasting of tropical cyclone activity over the western north pacific and the south china sea. *Weather and Forecasting*, 13(4):997 – 1004.
- Chand, S. S., Walsh, K. J. E., and Chan, J. C. L. (2010). A bayesian regression approach to seasonal prediction of tropical cyclones affecting the fiji region. *Journal of Climate*, 23(13):3425 – 3445.
- Chen, T., Fox, E., and Guestrin, C. (2014). Stochastic gradient hamiltonian monte carlo. In Xing, E. P. and Jebara, T., editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1683–1691, Beijing, China. PMLR.
- Choi, K.-S., Moon, J.-Y., Kim, D.-W., and Chu, P.-S. (2010). Seasonal prediction of tropical cyclone genesis frequency over the western north pacific using teleconnection patterns. *Theoretical and Applied Climatology*, 100:191–206.
- Fan, K. (2007). New predictors and a new prediction model for the typhoon frequency over western north pacific. *Science in China Series D: Earth Sciences*, 50(8):1417 – 1423.
- Fan, K. and Wang, H. (2009). A new approach to forecasting typhoon frequency over the western north pacific. *Weather and Forecasting*, 24(4):974 – 986.
- Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued garch process. *Journal of Time Series Analysis*, 27(6):923–942.
- Fokianos, K. (2012). 12 - count time series models. In Subba Rao, T., Subba Rao, S., and Rao, C., editors, *Time Series Analysis: Methods and Applications*, volume 30 of *Handbook of Statistics*, pages 315–347. Elsevier.
- Fokianos, K. and Fried, R. (2012). Interventions in log-linear poisson autoregression. *Statistical Modelling*, 12(4):299–322.
- Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439.

- Fokianos, K. and Tjøstheim, D. (2011). Log-linear poisson autoregression. *Journal of Multivariate Analysis*, 102(3):563–578.
- Friedman, J. H. (1985). Classification and multiple regression through projection pursuit. *Dept. of Statistics Tech. Rep. 12*.
- Ge, H., Xu, K., and Ghahramani, Z. (2018). Turing: a language for flexible probabilistic inference. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, pages 1682–1690.
- Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457 – 472.
- Griewank, A. and Walther, A. (2008). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, USA.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.
- Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(47):1593–1623.
- Hyndman, R. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts, Melbourne, Australia.
- Kendall, A. and Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, page 5580–5590, Red Hook, NY, USA. Curran Associates Inc.
- Lagmay, A. M. F., Agaton, R. P., Bahala, M. A. C., Briones, J. B. L. T., Cabacaba, K. M. C., Caro, C. V. C., Dasallas, L. L., Gonzalo, L. A. L., Ladiero, C. N., Lapidez, J. P., Mungcal, M. T. F., Puno, J. V. R., Ramos, M. M. A. C., Santiago, J., Suarez, J. K., and Tablazon, J. P. (2015). Devastating storm surges of typhoon haiyan. *International Journal of Disaster Risk Reduction*, 11:1–12.
- Lander, M. A. (1994). An exploratory analysis of the relationship between tropical storm formation in the western north pacific and enso. *Monthly Weather Review*, 122(4):636 – 651.
- Li, X., Yang, S., Wang, H., Jia, X., and Kumar, A. (2013). A dynamical-statistical forecast model for the annual frequency of western pacific tropical cyclones based on the ncep climate forecast system version 2. *Journal of Geophysical Research: Atmospheres*, 118(21):12,061–12,074.
- Magee, A. D., Kiem, A. S., and Chan, J. C. L. (2021). A new approach for location-specific seasonal outlooks of typhoon and super typhoon frequency across the western north pacific region. *Scientific Reports*, 11:19439.
- McDonnell, K. A. and Holbrook, N. J. (2004a). A poisson regression model approach to predicting tropical cyclogenesis in the australian/southwest pacific ocean region using the soi and saturated equivalent potential temperature gradient as predictors. *Geophysical Research Letters*, 31(20).
- McDonnell, K. A. and Holbrook, N. J. (2004b). A poisson regression model of tropical cyclogenesis for the australian–southwest pacific ocean region. *Weather and Forecasting*, 19(2):440 – 455.
- McKenzie, E. (1985). Some simple models for discrete variate time series1. *JAWRA Journal of the American Water Resources Association*, 21(4):645–650.
- Murphy, A. H. (1992). Climatology, persistence, and their linear combination as standards of reference in skill scores. *Weather and Forecasting*, 7(4):692 – 698.
- Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3):705 – 767.
- Neal, R. M. (2011). Mcmc using hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*, pages 113–162. Chapman and Hall/CRC.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Seriño, M. N. V., Caverio, J. A., Cuizon, J., Ratilla, T. C., Ramoneda, B. M., Bellezas, M. H. I., and Ceniza, M. J. C. (2021). Impact of the 2013 super typhoon haiyan on the livelihood of small-scale

- coconut farmers in leyte island, philippines. *International Journal of Disaster Risk Reduction*, 52:101939.
- Tian, B. and Fan, K. (2019). Seasonal climate prediction models for the number of landfalling tropical cyclones in china. *Journal of Meteorological Research*, 33:837–850.
- Wang, B. and Chan, J. C. L. (2002). How strong enso events affect tropical storm activity over the western north pacific. *Journal of Climate*, 15(13):1643 – 1658.
- Weiss, C. H. (2018). *An Introduction to Discrete-Valued Time Series*. John Wiley & Sons Ltd, New Jersey, USA.
- Werner, A. and Holbrook, N. J. (2011). A bayesian forecast model of australian region tropical cyclone formation. *Journal of Climate*, 24(23):6114 – 6131.
- Zhang, W., Vecchi, G. A., Villarini, G., Murakami, H., Gudgel, R., and Yang, X. (2017). Statistical–dynamical seasonal forecast of western north pacific and east asia landfalling tropical cyclones using the gfdl flr coupled climate model. *Journal of Climate*, 30(6):2209 – 2232.
- Zhang, X., Zhong, S., Wu, Z., and Li, Y. (2018). Seasonal prediction of the typhoon genesis frequency over the western north pacific with a poisson regression model. *Climate Dynamics*, 51:4585–4600.