Multi-step-ahead Cyclone Intensity Prediction with Bayesian Neural Networks

Chapter · August 2019

DOI: 10.1007/978-3-030-29911-8_22

2 authors:

Ratneel Vikash Deo
University of the South Pacific

Rohitash Chandra
UNSW Sydney

Multi-step-ahead Cyclone Intensity Prediction with Bayesian Neural Networks

Ratneel Deo[1,2](✉) and Rohitash Chandra[2,3]

[1] School of Computing, Information and Mathematical Sciences, University of the South Pacific, Suva, Fiji
deo.ratneel@gmail.com
[2] Centre for Translational Data Science, The University of Sydney, Sydney, Australia
c.rohitash@gmail.com
[3] School of Geosciences, The University of Sydney, Sydney, Australia

Abstract. The chaotic nature of cyclones makes track and wind-intensity prediction a challenging task. The complexity in attaining robust and accurate prediction increases with an increase of the prediction horizon. There is lack of robust uncertainty quantification in models that have been used for cyclone prediction problems. Bayesian inference provide a principled approach for quantifying uncertainties that arise from model and data, which is essential for prediction, particularly in the case of cyclones. In this paper, Bayesian neural networks are used for multi-step ahead time series prediction for cyclones in the South Pacific region. The results show promising prediction accuracy with uncertainty quantification for shorter prediction horizon; however, the challenge lies in higher prediction horizons.

1 Introduction

A tropical cyclone is a low pressure system with organized convection that forms over warm tropical waters [1]. Once formed, a cyclone can move over the ocean in the direction away from the equator that can last a few days to weeks [16]. The past decade has witnessed rise of destructive storms and cyclones with effect of climate change [22,35]. The wind-intensity and duration of cyclones are the major contributing factors of the extent of damages which can be greatly reduced with timely and precise warning [14,21,23]. The chaotic nature of cyclones makes prediction of track and wind-intensity a challenging task [36]. Statistical models have been a standard approach for predicting the intensity of the cyclones [24]. A popular method was the statistical hurricane intensity prediction scheme (SHIPS) [17] that made cyclone intensity forecasts for 12 to 120 h. SHIPS has been unsuitable for cyclones near the coast as the model was developed using cyclone data away from coastlines [18].

Machine learning methods have established themselves as complementary approaches for tropical cyclone tracking and prediction [4,13,25,27]. Amongst machine learning methods, neural networks have shown great promise due to

A. C. Nayak and A. Sharma (Eds.): PRICAI 2019, LNAI 11671, pp. 282–295, 2019.
https://doi.org/10.1007/978-3-030-29911-8_22

their prediction accuracy [6,8,20]. Baik and Huwang [3] presented of the earliest works where neural networks for tropical cyclone intensity prediction in the western North Pacific ocean for selected time-spans (12 to 72 h). Liu and Feng [30] used neural networks the prediction of the maximum potential intensity of cyclones. Chaudhuri *et al.* [13] considered tropical cyclones over the Bay of Bengal and the Arabian Sea for prediction of track and wind-intensity using feedforward neural networks. The method gave comparable performance with existing numerical models for 6 hour-ahead forecasts. Chandra and Dayal [8] presented a method for cyclone track prediction using coevolutionary Elman recurrent neural networks (RNNs) for the South Pacific ocean which was later improved [11]. Coevolutionary Elman RNNs was also used for cyclone wind-intensity prediction [7]. Deo and Chandra [19] presented the problem of predicting rapid intensification in wind-intensity of tropical cyclones with Elman RNNs. More recently, Zhang *et al.* [37] used matrix neural networks for prediction of tracks for South Indian and South Pacific oceans and found that their method performed better when compared to prominent deep learning architectures such as long-short term memory networks (LSTMs) and gated recurrent units (GRUs).

Bayesian inference accounts for uncertainties in parameter estimates and propagates it into predictions by marginalizing them out of the posterior distribution [32]. Bayesian neural networks (BNNs) use Markov Chain Monti-Carlo (MCMC) methods to sample from the posterior distribution of weights and biases [31]. MCMC methods have limitations due to the lack of scalability given the increasing size of the model and data which has been addressed by combining them with the gradient-based learning methods [10,28], simulated annealing [29], and evolutionary algorithms [26]. Bayesian methods have been used for cyclone identification problems in the past and have shown promising results [5,15].

There exists limited literature on the use of Bayesian neural methods for multiple step-ahead predictions. In the literature, cyclone wind-intensity prediction via neural networks has mostly been approached as a single-step ahead prediction problem [7,8,37]. Given dataset Joint Typhoon Warning Centre [2], this implies that single step-ahead prediction provides only a six-hour ahead prediction. This can be extended if the prediction is fed back as input to the model; however, this can reduce accuracy in prediction [33]. A limited prediction horizon lacks usefulness in giving a warning for evacuation and preparations in case the cyclone intensifies in the future. Therefore, it would be useful to increase the prediction horizon while featuring uncertainty quantification in the prediction via Bayesian inference. Moreover, there is limited work done using Bayesian inference in the area of storms and cyclones.

In this paper, Bayesian neural networks are used for multi-step ahead time series prediction for cyclones in the South Pacific region. This enables uncertainty quantification in the predictive model given multiple steps-ahead. We use two MCMC sampling methods and compare the performance with standard backpropagation network given increasing prediction horizons.

The rest of the paper is organised as follows. The proposed method is presented in Sect. 2 which is followed by results and discussion in Sect. 3. Section 4 concludes the paper with discussion of future work.

2 Bayesian Neural Networks for Cyclone Intensity Prediction

In this section, we present the Bayesian neural network (BNN) model used for multi-step ahead cyclone intensity prediction. The dynamics of a feedforward network is given in Eq. 1.

$$f(\mathbf{x}_t) = g\bigg(\delta_o + \sum_{h=1}^{H} v_j g\bigg(\delta_h + \sum_{d=1}^{D} w_{dh} y_{t-d})\bigg) \tag{1}$$

where δ_o and δ_h are the bias weights for the output o and hidden h layer, respectively. V_j is the weight which maps the hidden layer h to the output layer. w_{dh} is the weight which maps y_{t-d} to the hidden layer h, and g is the activation function given in Eq. 2. In our case, we use the sigmoid activation function in the hidden and output layer units.

$$f(y_i) = \frac{1}{1 + e^{-y_i}} \tag{2}$$

Taken's embedding theorem [34] is used to reconstruct the univariate time series data into a phase state as given in Eq. 3.

$$Y(t) = [(x(t), x(t-T), ..., x(t(D-1)T)] \tag{3}$$

where $x(t)$ is the observed time series, T is the time delay, D is the embedding dimension, $t = 0, 1, 2, ..., N - DT - 1$ and N is the length of the original time series.

2.1 Model and Priors

Let y_t denote a univariate time series of cyclone wind-intensity:

$$y_t = f(\mathbf{x}_t) + \epsilon_t, \text{ for } t = 1, 2, \ldots, n \tag{4}$$

where $f(\mathbf{x}_t) = E(y_t|\mathbf{x}_t)$, is an unknown function, $\mathbf{x}_t = (y_{t-1}, \ldots, y_{t-D})$ is a vector of lagged values of y_t, and ϵ_t is the noise with $\epsilon_t \sim \mathcal{N}(0, \tau^2)\ \forall t$.

Let $\boldsymbol{\theta} = (\tilde{\mathbf{w}}, \mathbf{v}, \boldsymbol{\delta}, \tau^2)$, with $\boldsymbol{\delta} = (\delta_o, \delta_h)$, denote $L = (DH + (2 * H) + O + 1)$ vector of parameters that includes weights and biases, with O number of neurons in output layer. H is the number of hidden neurons required to evaluate the likelihood for the model given by Eq. 4, with $\tilde{\mathbf{w}} = (\mathbf{w_{1.}}', \ldots, \mathbf{w}'_{D.})'$, and $\mathbf{w_{d.}} = (w_{d1}, \ldots, w_{dH})'$, for $d = 1, \ldots, D$.

Bayesian inference via MCMC sampling is implemented by drawing samples from the posterior distribution given the prior distribution and the likelihood function that verifies how well the model fits the data. Given our model is the feedforward network given in Eq. 1, the prior distributions for the elements of $\boldsymbol{\theta}$ are given as

$$\begin{aligned}
v_h &\sim \mathcal{N}(0, \sigma^2) \text{ for } h = 1, \ldots, H, \\
\delta_0 &\sim N(0, \sigma^2) \\
\delta_h &\sim N(0, \sigma^2) \\
w_{dj} &\sim \mathcal{N}(0, \sigma^2) \text{ for } h = 1, \ldots, H \text{ and } d = 1, \ldots, D, \\
\tau^2 &\sim \mathcal{IG}(\nu_1, \nu_2)
\end{aligned} \tag{5}$$

where H is the number of hidden neurons. In general, the log posterior is

$$\log\left(p(\boldsymbol{\theta}|\mathbf{y})\right) = \log\left(p(\boldsymbol{\theta})\right) + \log\left(p(\mathbf{y}|\boldsymbol{\theta})\right)$$

In our particular model, the log likelihood is

$$\begin{aligned}
\log\left(p(\mathbf{y}_{\mathcal{A}_{\mathbf{D},\mathbf{T}}}|\boldsymbol{\theta})\right) = &-\frac{n-1}{2}\log(\tau^2) \\
&-\frac{1}{2\tau^2}\sum\nolimits_{t\in\mathcal{A}_{\mathcal{D},\mathcal{T}}}\left(y_t - E(y_t|\mathbf{x}_t)\right)^2
\end{aligned} \tag{6}$$

where $E(y_t|\mathbf{x}_t)$ is given by Eq. 1. We further assume that the elements of $\boldsymbol{\theta}$ are independent *apriori*; therefore, the log of the prior distributions is

$$\begin{aligned}
\log\left(p(\boldsymbol{\theta})\right) = &-\frac{HD+H+2}{2}\log(\sigma^2) \\
&-\frac{1}{2\sigma^2}\left(\sum_{h=1}^{H}\sum_{d=1}^{D} w_{dh}^2 + \sum_{h=1}^{H}(\delta_h^2 + v_h^2) + \delta_o^2\right) \\
&-(1+\nu_1)\log(\tau^2) - \frac{\nu_2}{\tau^2}
\end{aligned} \tag{7}$$

where, σ^2, ν_1, ν_2 are user chosen constants. We feature MCMC random-walk and Langevin-gradient proposals for sampling [10]. In Langevin-gradient proposals, a stochastic noise is added to one step of gradients as shown in Eqs. 8 to 11.

$$\boldsymbol{\theta}^p \sim \mathcal{N}(\bar{\boldsymbol{\theta}}^{[k]}, \Sigma_\theta), \text{ where} \tag{8}$$

$$\bar{\boldsymbol{\theta}}^{[k]} = \boldsymbol{\theta}^{[k]} + r \times \nabla E_{\mathbf{y}_{\mathcal{A}_{\mathbf{D},\mathbf{T}}}}[\boldsymbol{\theta}^{[k]}], \tag{9}$$

$$E_{\mathbf{y}_{\mathcal{A}_{\mathbf{D},\mathbf{T}}}}[\boldsymbol{\theta}^{[k]}] = \sum_{t\in\mathcal{A}_{D,T}} (y_t - f(\mathbf{x}_t)^{[k]})^2, \tag{10}$$

$$\nabla E_{\mathbf{y}_{\mathcal{A}_{\mathbf{D},\mathbf{T}}}}[\boldsymbol{\theta}^{[k]}] = \left(\frac{\partial E}{\partial\theta_1}, \ldots, \frac{\partial E}{\partial\theta_L}\right) \tag{11}$$

where r is the learning rate, $\Sigma_\theta = \sigma_\theta^2 I_L$ and I_L is the $L \times L$ identity matrix. The newly proposed value of $\boldsymbol{\theta}^p$ consists of 2 parts:

1. A gradient descent based weight update given by Eqs. 8 to 11.
2. Add an amount of noise, from $\mathcal{N}(0, \Sigma_\theta)$.

2.2 Framework

The framework shown in Fig. 1 gives an overview of the Bayesian neural network architecture used for multi-step cyclone intensity prediction. We use a three-layer feedforward neural network and process data according to the prediction horizon using Taken's theorem.

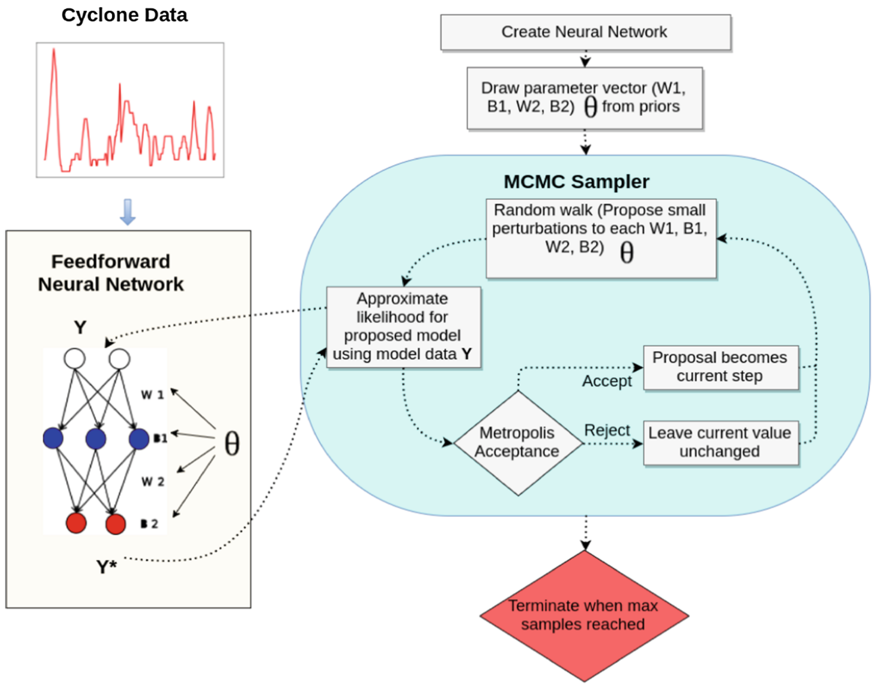


Fig. 1. Bayesian neural network for cyclone intensity prediction. The figure shows the mapping between the MCMC sampler and Bayesian neural network parameters. The three-layer feedforward neural network has four sets of parameters that make up $\boldsymbol{\theta}$; *W1* corresponds to the network weights between the input layer and the hidden layer, *B1* corresponds to the biases of each of the hidden neurons, and *W2, B2* correspond to the weights and biases of the output layer

Algorithm 1 begins by pre-processing the data using Taken's state space reconstruction described in Eq. 3. The neural network model is then created according the specified number of input, hidden and output neurons. Note that the number of output neurons defines the steps-ahead or prediction horizon. The sampling begins where a proposal for $\boldsymbol{\theta}$ that features the training neural network parameters (weights and biases) is generated either using a random-walk (RW-MCMC) or using Langevin-gradient (LG-MCMC) proposals as defined by the user. The algorithm calculates the log likelihood and then given the prior, it either accepts or rejects the proposals using the Metropolis-Hastings criterion. If the sample is accepted, it becomes part of the posterior distribution $\boldsymbol{\theta}$; otherwise, the previously accepted sample is retained. The process repeats until the maximum number of samples (max samples) is reached. Once this is done, the trained Bayesian neural network is tested using unseen data as defined by the user.

Alg. 1. Bayesian neural networks using MCMC sampling

Data: Univariate time series of cyclone intensities $\mathbf{y}$
Result: Posterior of weights and biases $p(\boldsymbol{\theta}|\mathbf{y})$
Step 1: State-space reconstruction by Equation 3
Step 2: Define feedforward network as given in Equation 1
Step 3: Define $\boldsymbol{\theta}$ as the set of all weights and biases
Step 4: Set parameters σ^2 ,ν_1, ν_2 for prior given in Equation 7

for *each k until max-samples* **do**
    1. Draw $\boldsymbol{\eta}$ from $\mathcal{N}(0, \Sigma_\eta)$
    2. Propose $\boldsymbol{\theta}^* = \theta^{[k]} + \boldsymbol{\eta}$
    Draw $l \sim U[0,1]$
    if *(LG is true) and ($l < LG - freq$)* **then**
        Conditional on m^c, propose a new value of θ
        using Langevin-gradient proposal distribution $\theta^p \sim q(\theta|\theta^c, m^c)$
        where, $q(\theta|\theta^c, m^c)$ is given in Equations 8 to 11
    else
        Conditional on m^c, propose a new value of θ
        using Random-Walk proposal distribution
    end
    3. Compute acceptance probability α
    if $u < \alpha$ **then**
        $\boldsymbol{\theta}^{[k+1]} = \boldsymbol{\theta}^*$
    end
    else
        $\boldsymbol{\theta}^{[k+1]} = \boldsymbol{\theta}^{[k]}$
    end
end

3 Experiments and Results

In this section, we present the experiments that are conducted to evaluate the efficiency of Bayesian neural networks for multi-step-ahead cyclone wind-intensity prediction.

3.1 Dataset and Experiment Design

The cyclone data featured the South Pacific ocean that considered cyclones from 1985 to 2013 taken from Joint Typhoon Warning Centre [2]. There are a total of 280 cyclones in the dataset with an average duration of 8 days and a standard deviation of 5.5 days for each cyclone. We concatenate the cyclones into a single time series by stacking the cyclone data in ascending order based on its start time as done in the literature [8,19]. Each cyclone was recorded at 6-h regular intervals, and therefore, the dataset consisted of 9,400 data points when concatenated. We implemented a 70:30 split ratio for our training and testing data. We normalize the univariate cyclone time series between 0 to 1 and use Taken's embedding theorem [34] from Eq. 3 to reconstruct the data where time lag of $T = 2$ and embedding dimension $D = 5$.

We use a multi-output neural network to make multiple-step ahead prediction based on the prediction horizon, where the number of outputs in each data instance represents the prediction horizon. Each output of the neural network corresponds to one step in the prediction horizon; i.e. the first output of the network maps on to the first and the second output represents the prediction horizon, and the rest in a similar manner. The root mean squared-error (RMSE) is used to measure the prediction performance of the neural network.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{12}$$

where $y_i, \hat{y}_i$ are the observed data and predicted data, respectively. N is the length of the observed data.

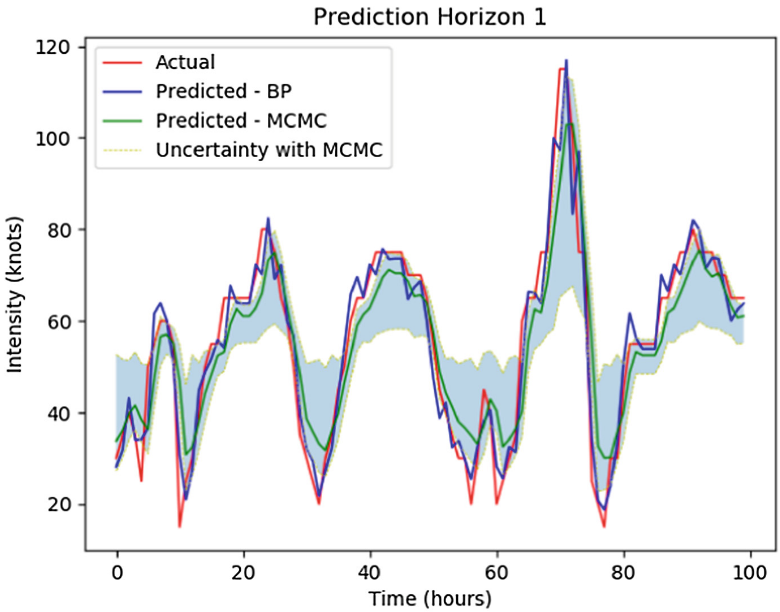
We implemented two experimental configurations with a different number of output neurons to enable predictions over two sets of the prediction horizon.

– Experiment 1: Five step-ahead prediction with 5 input neurons, 10 hidden neurons, and 5 output neurons
– Experiment 2: Ten step-ahead prediction with 5 input neurons, 10 hidden neurons, and 10 output neurons

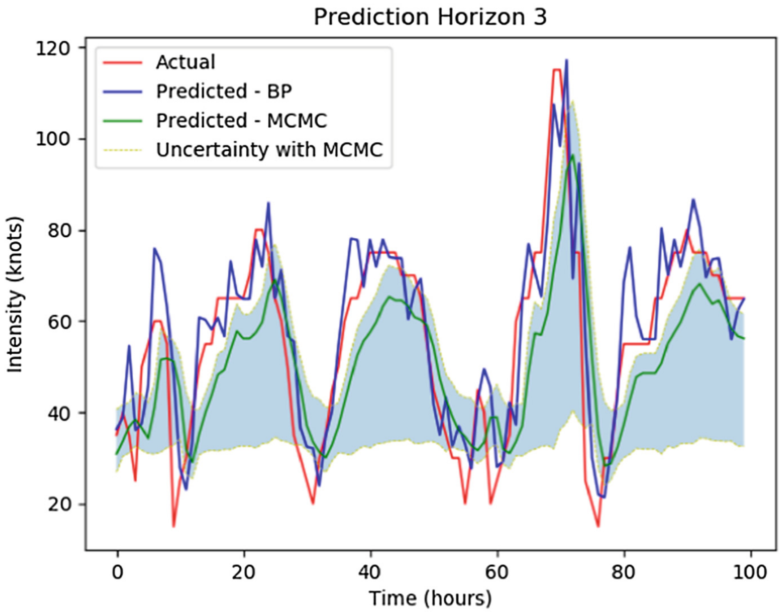
Bayesian neural networks with random-walk (RW-MCMC) and Langevin-gradient (LG-MCMC) sampling methods for forecasting cyclone wind-intensity with the above experimental setup. We also provide results for backpropagation neural network (BP) using gradient-descent training. In the case of RW-MCMC, each proposal of the Markov in the sampling chain (θ) added a Gaussian with mean $\mu = 0$ and standard deviation $\sigma = 0.02$. A Gaussian noise was also used for sampling η with mean $\mu = 0$ and standard deviation $\sigma = 0.2$. T The parameters for the prior given in Eq. 7 was set: $\sigma^2 = 25$, $\nu_1 = 0$, $\nu_2 = 0$. In the case of backpropagation, fixed learning rate of 0.1 and training time of 2000 epochs was used. A similar configuration was used with LG-MCMC; however, the additional parameter involved was the learning rate of 0.1 for Langevin-gradients. The maximum sampling time for the MCMC methods were set to 10,000 samples.

3.2 Results

Table 1 provides the results for the prediction accuracy for the respective methods. We observe that the three methods give a similar performance in most of the cases. There seems to be a linear relationship between the prediction horizon and prediction error; we notice a steady increase in the error in prediction

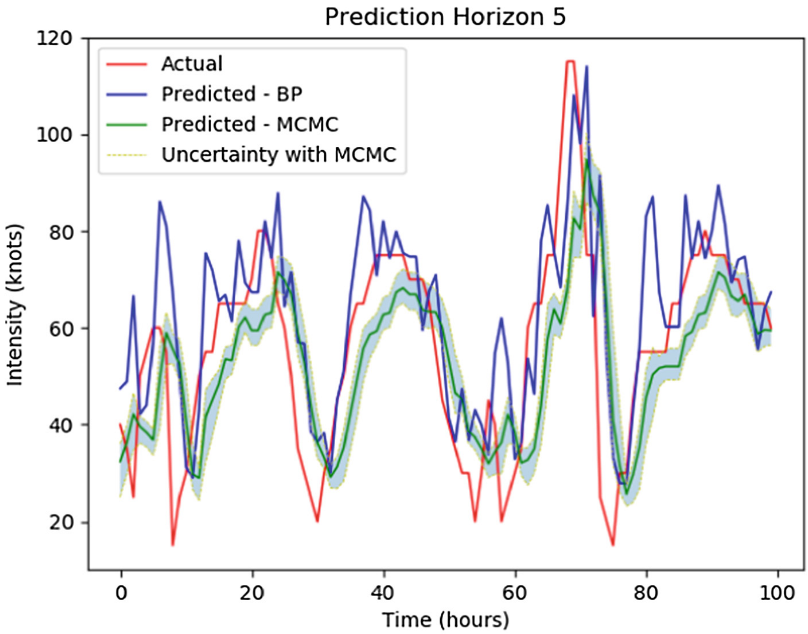


(a)

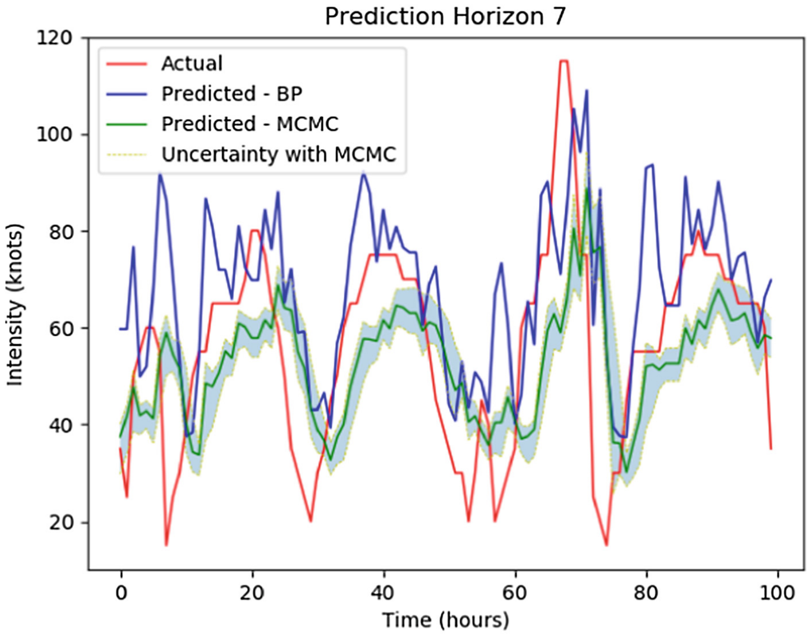


(b)

Fig. 2. (a) One step ahead prediction; (b) intensity prediction for third prediction horizon.

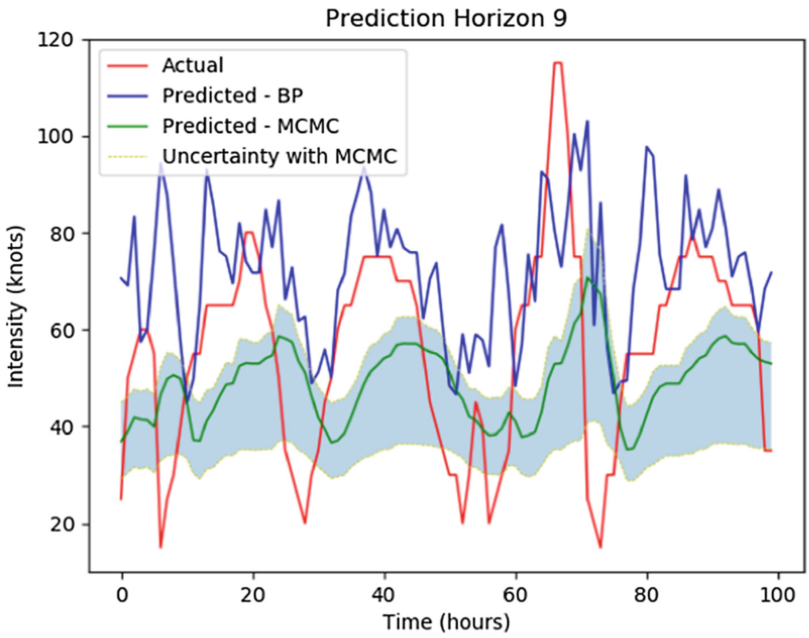


(a)

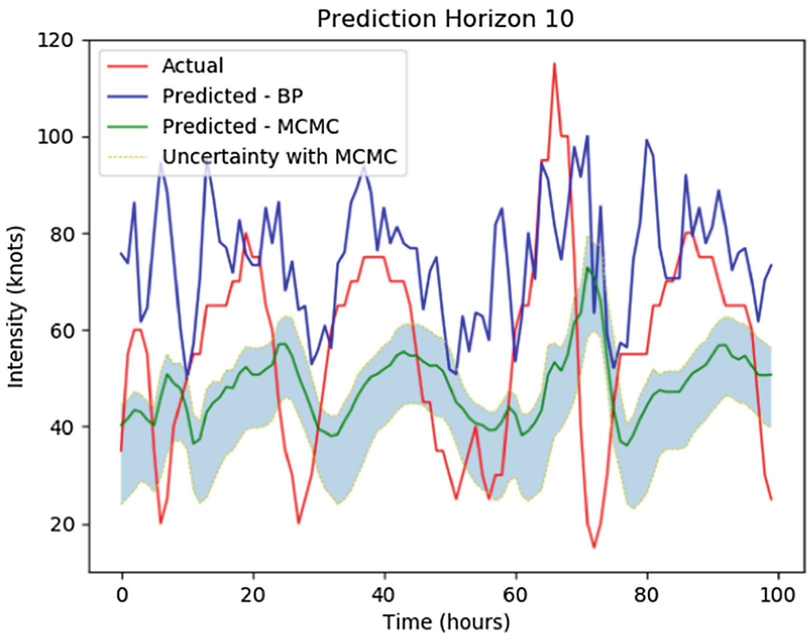


(b)

Fig. 3. (a) intensity prediction for fifth prediction horizon; (b) intensity prediction for seventh prediction horizon.



(a)



(b)

Fig. 4. (a) intensity prediction for ninth prediction horizon; (b) intensity prediction for tenth prediction horizon.

with increasing prediction horizon. This relationship is apparent in the case with backpropagation; however, MCMC methods do not show this relationship as clearly.

We provide a vitalisation of the prediction performance and notice a similar trend in Figs. 2, 3 and 4 where the predictions deviate further from the actual data with increasing prediction horizon. We notice that the prediction given by backpropagation is more chaotic with increasing prediction horizons. As we move between different prediction horizons, the uncertainty in prediction given by RW-MCMC becomes very inconsistent. There is an increase in uncertainty moving from horizon 1 to 3; however, there is a decrease in uncertainty while moving from horizon 3 to 5 as shown in Figs. 2(a), (b), and 3(a), respectively. The prediction horizon of 5 to 7 in Figs. 3(a) and (b) show less uncertainty in prediction; however, the prediction accuracy is poor. It is not surprising that, the poorest performance in terms of accuracy is seen with the highest prediction horizons. The figures do not include prediction given by LD-MCMC as the results were almost identical to results produced by RW-MCMC.

Table 1. Prediction error on testing data over 10 time-steps

Step	BP	RW-MCMC	LG-MCMC
1	0.031 ± 0.0005	0.036 ± 0.0057	0.034 ± 0.0012
2	0.044 ± 0.0005	0.044 ± 0.0061	0.045 ± 0.0007
3	0.055 ± 0.0008	0.053 ± 0.0118	0.065 ± 0.0021
4	0.078 ± 0.0011	0.067 ± 0.0238	0.079 ± 0.0018
5	0.094 ± 0.0015	0.086 ± 0.0422	0.094 ± 0.0017
6	0.112 ± 0.0022	0.101 ± 0.0442	0.102 ± 0.0018
7	0.128 ± 0.0030	0.098 ± 0.0143	0.121 ± 0.0023
8	0.145 ± 0.0041	0.115 ± 0.0309	0.132 ± 0.0030
9	0.157 ± 0.0050	0.127 ± 0.0229	0.141 ± 0.0027
10	0.173 ± 0.0062	0.128 ± 0.0170	0.147 ± 0.0027

3.3 Discussion

The results, in general, have shown that backpropagation and Bayesian neural networks deteriorating performance for cyclone intensity over longer prediction horizons. Backpropagation neural network that used gradient descent has given a notable performance in terms of prediction accuracy. It showed consistency in the results where a clear relationship between prediction horizon and prediction error was noted. This behaviour was expected as there is increased complexities of predicting longer prediction horizons of the time series. Bayesian neural networks with two sampling methods showed similar relationship however, with slight deviations. The general trend of increased prediction error with increasing horizons are seen; however, there are some outliers in the trend given the uncertainties in prediction.

The larger prediction horizons are vital. Since cyclone intensity determines the extent of damages; the better our prediction for longer prediction horizons, the more time the general public get to prepare. The 10-step ahead prediction would give 2.25 days of advance warning which is ample time to prepare or evacuate safely [14,23]. We found that increasing prediction horizons increases the complexity of the problem. This could be addressed in future work by using more advanced MCMC methods such as parallel tempering [12]. Another limitation of the experiments is in the neural network model which requires 5 data points, in order to make a prediction. This implies 1.25 days of cyclone activity before the model could be used. The limitation can be addressed in future by using a dynamic time series prediction neural network model [9]. The dynamic neural network would be able begin prediction without waiting for 5 data points.

4 Conclusions and Future Work

We have presented a Bayesian neural network approach to the prediction of wind-intensity of cyclones given multiple prediction horizons. The results show that the proposed method achieves more accurate predictions for shorter prediction horizon when compared to longer ones, which is not surprising as it is well known that the complexity of the problem increases with the horizon. Nevertheless, the proposed method has the power of rigorous uncertainty quantification that can be propagated for different prediction horizons which is helpful for extreme weather forecasting.

In future work, Bayesian neural networks can be further improved by incorporating more advanced models such as recurrent neural networks and other deep learning architectures. More datasets and input features such as distance to landfall, sea-surface temperature, and humidity can help in achieving better predictions. Furthermore, the approach can be used for trajectory or path prediction of cyclones as it provides uncertainty quantification in prediction.

Supplementary Material

The software code and dataset is given online[1].

Acknowledgements. We would like to acknowledge Prof. Sally Cripps for discussion and support during the course of this project.

References

1. Bureau of meteorology research centre (2000). http://www.bom.gov.au/cyclone/about/. Accessed 02 Jan 2018
2. JTWC tropical cyclone best track data site (2018). http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/
3. Baik, J.J., Hwang, H.S.: Tropical cyclone intensity prediction using regression method and neural network. J. Meteorol. Soc. Jpn. Ser. II **76**(5), 711–717 (1998)

[1] https://github.com/sydney-machine-learning/Bayesian-neural-networks-cyclone-intensity.

4. Carr III, L.E., Elsberry, R.L., Peak, J.E.: Beta test of the systematic approach expert system prototype as a tropical cyclone track forecasting aid. Weather Forecast. **16**(3), 355–368 (2001)
5. Chand, S.S., Walsh, K.J., Chan, J.C.: A Bayesian regression approach to seasonal prediction of tropical cyclones affecting the Fiji region. J. Clim. **23**(13), 3425–3445 (2010)
6. Chandra, R.: Multi-objective cooperative neuro-evolution of recurrent neural networks for time series prediction. In: IEEE Congress on Evolutionary Computation, Sendai, Japan, pp. 101–108, May 2015
7. Chandra, R., Dayal, K.: Cooperative coevolution of Elman recurrent networks for tropical cyclone wind-intensity prediction in the South Pacific region. In: IEEE Congress on Evolutionary Computation, Sendai, Japan, pp. 1784–1791, May 2015
8. Chandra, R., Dayal, K., Rollings, N.: Application of cooperative neuro-evolution of Elman recurrent networks for a two-dimensional cyclone track prediction for the South Pacific region. In: International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, pp. 721–728, July 2015
9. Chandra, R.: Dynamic cyclone wind-intensity prediction using co-evolutionary multi-task learning. In: Liu, D., Xie, S., Li, Y., Zhao, D., El-Alfy, E.-S.M. (eds.) ICONIP 2017. LNCS, vol. 10638, pp. 618–627. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70139-4_63
10. Chandra, R., Azizi, L., Cripps, S.: Bayesian neural learning via Langevin dynamics for chaotic time series prediction. In: Liu, D., Xie, S., Li, Y., Zhao, D., El-Alfy, E.-S.M. (eds.) ICONIP 2017. LNCS, vol. 10638, pp. 564–573. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70139-4_57
11. Chandra, R., Deo, R., Omlin, C.W.: An architecture for encoding two-dimensional cyclone track prediction problem in coevolutionary recurrent neural networks. In: 2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, pp. 4865–4872, 24–29 July 2016
12. Chandra, R., Jain, K., Deo, R.V., Cripps, S.: Langevin-gradient parallel tempering for Bayesian neural learning. Neurocomputing (2019). http://www.sciencedirect.com/science/article/pii/S0925231219308069
13. Chaudhuri, S., Dutta, D., Goswami, S., Middey, A.: Track and intensity forecast of tropical cyclones over the North Indian Ocean with multilayer feed forward neural nets. Meteorol. Appl. **22**(3), 563–575 (2015)
14. Chowdhury, A.M.R., Bhuyia, A.U., Choudhury, A.Y., Sen, R.: The Bangladesh cyclone of 1991: why so many people died. Disasters **17**(4), 291–304 (1993)
15. Chu, P.S., Zhao, X.: A Bayesian regression approach for predicting seasonal tropical cyclone activity over the central North Pacific. J. Clim. **20**(15), 4002–4013 (2007)
16. Debsarma, S.: Cyclone and its warning system in Bangladesh. National Disaster Reduction Day (2001)
17. DeMaria, M., Kaplan, J.: A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. Weather Forecast. **9**, 209–220 (1994)
18. DeMaria, M., Mainelli, M., Shay, L., Knaff, J., Kaplan, J.: Futher improvements to the statistical hurricane intensity prediction scheme (SHIPS). Weather Forecast. **20**, 531–543 (2005)
19. Deo, R., Chandra, R.: Identification of minimal timespan problem for recurrent neural networks with application to cyclone wind-intensity prediction. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 489–496. IEEE (2016)

20. Du, W., Leung, S.Y.S., Kwong, C.K.: Time series forecasting by neural networks: a knee point-based multiobjective evolutionary algorithm approach. Expert Syst. Appl. **41**(18), 8049–8061 (2014)
21. Dvorak, V.F.: Tropical cyclone intensity analysis using satellite data, vol. 11. US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service (1984)
22. Handmer, J., Iveson, H., et al.: Cyclone Pam in Vanuatu: learning from the low death toll. Aust. J. Emerg. Manag. **32**(2), 60 (2017)
23. Haque, C.E.: Atmospheric hazards preparedness in Bangladesh: a study of warning, adjustments and recovery from the April 1991 cyclone. Earthq. Atmos. Hazards **16**, 181–202 (1997)
24. Jarvinen, B.R., Neumann, C.J.: Statistical forecasts of tropical cyclone intensity. NOAA Techonological Memorandum. NWS NHC-10, p. 22 (1979)
25. Jin, L., Yao, C., Huang, X.Y.: A nonlinear artificial intelligence ensemble prediction model for typhoon intensity. Mon. Weather Rev. **136**, 4541–4554 (2008)
26. Kocadağlı, O., Aşıkgil, B.: Nonlinear time series forecasting with Bayesian neural networks. Expert Syst. Appl. **41**(15), 6596–6610 (2014)
27. Kovordányi, R., Roy, C.: Cyclone track forecasting based on satellite images using artificial neural networks. ISPRS J. Photogramm. Remote Sens. **64**, 513–521 (2009)
28. Li, C., Chen, C., Carlson, D., Carin, L.: Preconditioned stochastic gradient Langevin dynamics for deep neural networks. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
29. Liang, F.: Annealing stochastic approximation Monte Carlo algorithm for neural network training. Mach. Learn. **68**(3), 201–233 (2007)
30. Liu, J., Feng, B.: A neural network regression model for tropical cyclone forecast. In: Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, pp. 4122–4128, 18–21 August 2005
31. MacKay, D.J.: A practical Bayesian framework for backpropagation networks. Neural Comput. **4**(3), 448–472 (1992)
32. MacKay, D.J.: Hyperparameters: optimize, or integrate out? In: Heidbreder, G.R. (ed.) Maximum Entropy and Bayesian Methods. Fundamental Theories of Physics (An International Book Series on The Fundamental Theories of Physics: Their Clarification, Development and Application), vol. 62, pp. 43–59. Springer, Dordrecht (1996). https://doi.org/10.1007/978-94-015-8729-7_2
33. Taieb, S.B., Sorjamaa, A., Bontempi, G.: Multiple-output modeling for multi-step-ahead time series forecasting. Neurocomputing **73**(10), 1950–1957 (2010)
34. Takens, F.: Detecting strange attractors in turbulence. In: Rand, D., Young, L.-S. (eds.) Dynamical Systems and Turbulence, Warwick 1980. LNM, vol. 898, pp. 366–381. Springer, Heidelberg (1981). https://doi.org/10.1007/BFb0091924
35. Tatham, P., Ball, C.M., Wu, Y., Diplas, P.: Using long endurance remotely piloted aircraft systems to support humanitarian logistic operations: a case study of cyclone winston. In: Smart Technologies for Emergency Response and Disaster Management, pp. 264–278. IGI Global (2018)
36. Zhang, F., Tao, D.: Effects of vertical wind shear on the predictability of tropical cyclones. J. Atmos. Sci. **70**(3), 975–983 (2013)
37. Zhang, Y., Chandra, R., Gao, J.: Cyclone track prediction with matrix neural networks. In: 2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, pp. 1–8, 8–13 July 2018