ISSN 2229-5518

Forecasting Daily Temperature in The City of Bandung, Indonesia by Means Bayesian Structural Time Series

I Gede Nyoman Mindra Jaya[1], Velya Novianty Putri[2]
[1, 2] Department Statistics of Padjadjaran University, Bandung, Indonesia

Abstract— Forecasting climate data is still become a crucial issue. The forecast reliability and accuracy is dependent on several factors such as data availability. In the era of big data, researchers tend to formulate algorithms that work well on a big data such as boosting methods. However, in developing countries, especially Indonesia, the availability of the complete climate data such as temperature still a crucial problem. A appropriate techniques is needed to forcast a small data set. Bayesian methods is already known as the statisticall methods which is work well for small data. The Bayesian structural time series was applied to forecast daily temperature data in the city of Bandung, Indonesia. We evaluated the model with three components are the trend, autoregressive and seasonal and compared to the ARIMA model. The model with a seasonal period equal to 90 was found as the best model based on testing and training method. It has the smallest root mean square error. The component trend is much higher influence the temporal pattern of temporal data compare than components autoregressive and seasonal. We approximated that the temperature in the city of Bandung during 2019 around 23.61 ^{0}C daily and the minimum and maximum temperatures are 22.54 ^{0}C and 24.43 ^{0}C respectively.

Index Terms— ARIMA, Autoregressive, Big data, Small data, Bayesian, Seasonal, Trend, Temperature

1 Introduction

CHILIMATE influences all species on earth and our daily life. Climate change may affect crop growth, transportation, business, ecology proces and many things else ([1], [2]). The one of climate variabels is Temperature. Forecasting climate data is still a crucial issue. The forecast reliability and accuracy is dependent on several factors such as data availability.

In the era of big data, researchers tend to formulate algorithms that work well on a big data such as boosting methods. However, in developing countries, especially Indonesia, the availability of the complete climate data such as temperature still a crucial problem. It means, the long data set of climate data are diffult to accessed. Thus, we are dealing with a case of small data set. The data obtained is not long enough to make good forecasting using classical statistical approaches such as autoregressive integrated moving average (ARIMA) or more modern approaches such as machine learning. However, these limitations must still be sought for a solution. One alternative method that can be used is to use the Bayesian approach [3]. This approach has long been recognized as an appropriate method for modeling small data [4]. This is because the Bayesian method utilizes information outside the research data that is better known as prior information. Combining prior information with information obtained from a limited data set is considered capable of increasing the level of precision of parameter estimation by minimizing error variance [5]. Bayesian structural time series models can be an alternative solution in modeling time series data for short data.

We apply the Bayesian structural time series modelling to model and forecast temperature data in the city of Bandung, Indonesi. Bandung is a capital city of West Java. Bandung is known as the tourist city. Many tourists come to visit each year. The daily temperature in Bandung vary over time. Planning visit time is influenced by climate condition such as the daily temperature of Bandung city. It is important to obtain a reliable forecast of temperature in Bandung to support Bandung as a city destination of tourism.

The paper is structured as follows: in Section 2, we define data and methodology, move on with result and discussion in Section 3. Finally, Section 4 concludes the paper with final remarks.

2 Method

Several time series method for foracasting purpose have been developed, however, there is no the best mothod for all condition. Here we introduced, the Bayesian structured time series method for modeling and forecasting temperature data in city of Bandung.

ISSN 2229-5518

2.1 Bayesian time series

The last few decades the Bayesian method has developed very rapidly. Bayesian techniques are very useful for modeling small datasets ([4], [5], [6]). Bayesian method was applied to time series analysis. Time series is defined as squence of data. Bayesian time series is also known as the probablitic time series $y_{1:T} = y_1, \dots, y_T$ require the spesification of joint distribution $p(y_{1:T})$. In time series analysis, generally assumed tha the $y_{1:T}$ are dependent. It is not feasible to assume that the time series are independent squaence. Thus modeling time series data is more difficult than indepndent corssection data with the likelihood function can be easily defined. To introduce statistica independence in time series data we need to develop a time steps. One procedure to introduce statistical independence is to use the Bayesian concept ([3], [7]:

$$p(y_T|y_{1:T-1}) = \frac{p(y_T, y_{1:T-1})}{p(y_{1:T-1})} = \frac{p(y_{1:T})}{p(y_{1:T-1})} \qquad (1)$$

By arranging (1) we obtain:

$$p(y_{1:T}) = p(y_T|y_{1:T-1})p(y_{1:T-1}) \qquad (2)$$

and we can express the joint probablity as:

$$p(y_{1:T}) = \prod_{t=1}^{T} p(y_t|y_{1:t-1}) \qquad (3)$$

$p(y_t|y_{1:t-1}) = p(y_t|y_{t-m:t-1})$ and it leads to the mth-order Markov model.

2.2 Structural time series

Let y_t denotes observation t in a real-valued of time series. A structural time series model can be defined by connecting y_t to a vector of latent state variables α_t:

$$y_t = Z_t^T \alpha_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, H_t) \qquad (4)$$

$$\alpha_{t+1} = \alpha_t + R_t \eta_t \qquad \eta_t \sim (0, Q_t) \qquad (5)$$

We call equations (4) as an observation equation and (5) as a transition equation. The matrix $Z_t, T_t,$ and $R_t$contain a mix of known values, and unknown parameters [8]. Model (4) and (5) describe a state space form. These models are generalization of ARIMA and VARIMA models. State space model can be extended to take account the time series components, trend, seasonality, regression effects and other state component that may be necessary. The basic state space model that we call as “basic structured model” can be written as:

$$\begin{aligned} y_t &= \mu_t + \tau_t + \beta' x_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \delta_{t-1} + u_t \\ \delta_t &= \delta_{t-1} + v_t \\ \tau_t &= -\sum_{s=1}^{S-1} \tau_{t-s} + w_t \end{aligned} \qquad (6)$$

where $\eta_t = (\varepsilon_t, u_t, w_t)$ in (6) denote the independent components of Gaussian random noise. The matrix $Q_t$is a constant diagonal matrix with diagonal elements σ_u^2, σ_v^2, and σ_w^2, and H_t is a constant σ_ε^2. This model accommodates the trend (δ_t), seasonal component (τ_t) the regression coefficients β with current level trend (intercept) is denoted by μ_t. Bayesian data augmentation is the primary tools that can be used to estimate the state space model. It works by producing simulations from posterior distribution $p(\alpha|y)$ [8].

2.3. Model evaluation

Usually in time series modelling, we estiamtes several parameters with different number components or different ordo of seasonal component. In order to compare the performance of those models we use training and testing method with several criteria will be used. The criater are root means squared deviance (RMSED) and mean absolute percentage deviance (MAPE) which are formulated as [9]:

$$RMSED = \sqrt{\frac{1}{H-1} \sum_{h=1}^{H} (y_{t+h} - \hat{y}_{t+h})^2} \qquad (7)$$

$$MAPE = \frac{1}{H} \left(\sum_{h=1}^{H} \left| \frac{y_{t+h} - \hat{y}_{t+h}}{y_{t+h}} \right| \right) 100\% \qquad (8)$$

with H is length of forecast and $\hat{y}$ denotes the forecast value. The smallest value RMSED, and MAPE indicates the best model. All the computation process will be done by R-software using package `forecast [10]` and `bsts [11]`.

3 Results

3.1 Data collecting

Bandung is the capital city and is located in the north central part of West Java province. Our dataset consists of daily temperatures from Bandung weather (LAPAN, latitude: 06° 53' 00,8" LS, longitude: 107° 35' 50,4"

ISSN 2229-5518

BT). We collect the data from LAPAN data sharing service system over the period of 27 September 2018 to 31 December 2018 with 461 total observations.

Before proceeding to detailed modeling and forecasting results, it is useful to get an overall feel for the daily average temperature data. Figure 2 shows the daily average temperature of Bandung from 27 September 2018 to 31 December 2018. The plot of temperature data is presented in Figure 1. Figure 1 suggests that there are trend and seasonal components in the temperature data of the city of Bandung, Indonesia.

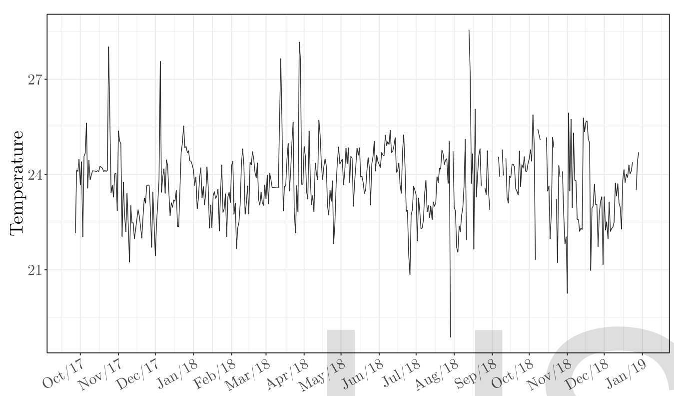

Figure 1. Temperatur data in the city of Bandung, September 2017 – December, 2018

3.2 Time series modelling

We developed six different models including ARIMA, Bayesian structural time series with trend and seasonal components. We evaluate the seasonal with $S = \{7, 30, 60, 90, 120\}$. Index $S = 7$ indicates the seasonal pattern is found for every one week and $S = 30, 60, 90,$ and 120 for one, two, three and four months. We use 300 first data set as training data set and 161 last data set as testing data. ARIMA model was estimated using `auto.arima` function in `forecast` package of R. Figure 2 presents the testing data model based on traning and testing validation method.

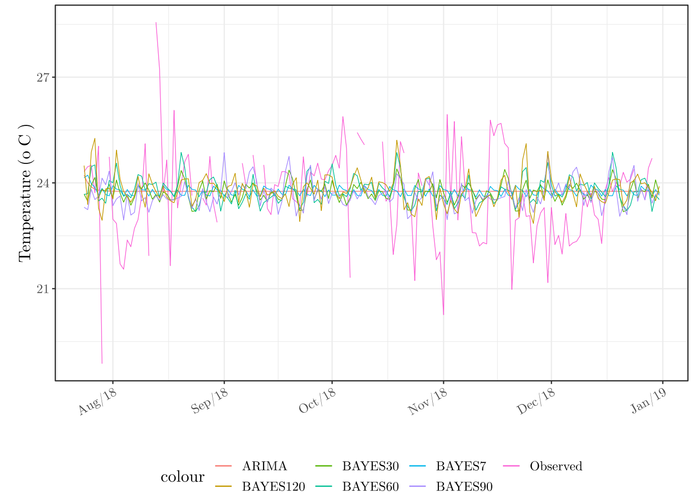

Figure 2. Model evaluation of ARIMA and Bayes structural time series with trend, autoregressive, and seasonal component for $S = \{7, 30, 60, 90, 120\}$

The ARIMA model produced a constant value of the forecast. It only differs for two first value of the forecast. It indicates ARIMA model cannot figure out the temporal trend and seasonal components in the temperature data. The best model that near to training data is the Bayesian structural time series data with $S = 90$. It is indicated by table 1 below that the BAYES S=90 has the smallest RMSED.

Table 1. Model evaluation

Model	RMSED	MAPE
ARIMA	0.784	0.020
BAYES S=7	0.726	0.020
BAYES S=30	0.544	0.020
BAYES S=60	1.069	0.021
BAYES S=90	0.097	0.020
BAYES S=120	0.945	0.020

3.3 Forecasting

Model Bayesian structural time series with three components: autoregressive, trend, and seasonal with $S = 90$ was selected as the best model based on RMSED criteria.

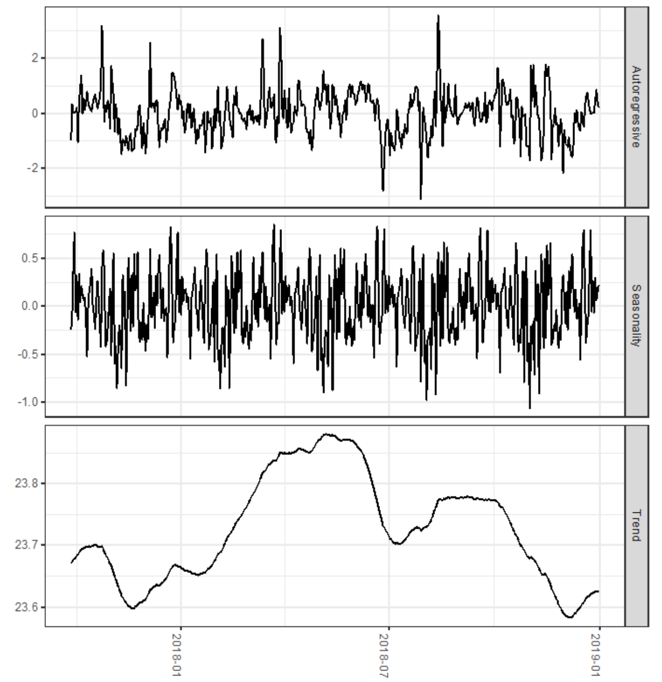

Figure 3. Autoregressive, Seasonal and Trend components of temperature data

Seasonal index $S = 90$ indicates that the temporal pattern of temperature is repeated for every 3 months. Based on this selected model, the forecasting value during 1 January to December 2019 was obtained and presented in Figure 4.

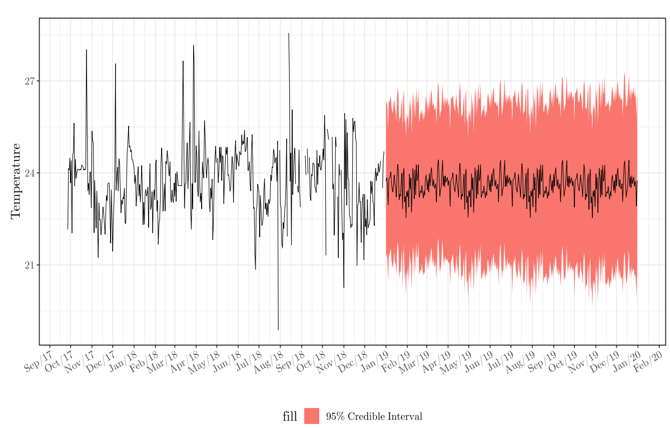

Figure 4. Forecasting temperature, 1 January- 31 December, 2019

Figure 4 shows that the temperature value during 1 January to December 2019. The temperature value has autoregressive, trend and seasonal pattern. It looks varies daily. However, the pattern is reapeted for every 3 months. The minimum values of the temperature is observed 22.54 ^{0}C and the maximum temperature is 24.43 ^{0}C with average 23.61 ^{0}C. The minimum temperature is observed during January to March and high temperature is observed during April to June.

4 CONCLUSION

The Bayesian structural time series was applied to forecast daily temperature data in the city of Bandung, Indonesia. We evaluated the model with three components are the trend, autoregressive and seasonal and compared to the ARIMA model. The model with a seasonal period equal to 90 was found as the best model based on testing and training method. It has the smallest root mean square error. The component trend is much higher influence the temporal pattern of temporal data compare than components autoregressive and seasonal. We approximated that the temperature in the city of Bandung during 2019 around 23.61 ^{0}C daily and the minimum and maximum temperatures are 22.54 ^{0}C and 24.43 ^{0}C respectively.

ACKNOWLEDGMENT

This paper is funded by the RFU Unpad contract: 1732 d/UN6.RKT/LT/2018. The authors thank Rector Universitas Padjadjaran. LAPAN Bandung for the temperature data and to the anonymous referee whose valuable checking has improved this paper

REFERENCES

[1] N. Wanders, M. T. H. v. Vliet, Y. Wada, M. F. P. Bierkens and L. P. H. (. v. Beek, "High-Resolution Global Water Temperature Modeling," *Water Resources Research*, vol. 1, no. 1, pp. 2760-2778, 2019.

[2] M. Andrade-Bejarano, "Monthly Average Temperature Modeling in an Intertropical Region," *American Meteorological Society*, vol. 1, no. 1, pp. 1099-1115, 2013.

[3] D. Barber, A. T. Cemgil and S. Chiappa, Bayesian Time Series Model, Cambridge: Cambridge University, 2011.

[4] M. Blangiardo and M. Cameletti, Spatial and Spatio-temporal Bayesian Models with R - INLA, Chichester: John Wiley & Sons, 2015.

[5] G. Koop, D. J. Poirier and J. L. Tobias, Bayesian Econometric Methods, New York: Cambridge Press, 2007.

[6] I. Jaya, H. Folmer, B. N. Ruchjana, F. Kristiani and A. Yudhie, "Modeling of Infectious Diseases: A Core Research Topic for the Next Hundred Years," in *Regional Research Frontiers - Vol. 2 Methodological Advances, Regional Systems Modeling and Open Sciences*, USA, Springer International Publishing, 2017, pp. 239-254.

[7] S. S. Rangapuram, M. Seeger, J. Gasthaus, L. Stella, Y. Wang and T. Januschowski, "Deep State Space Models for Time Series Forecasting," in *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Canada, 2018.

[8] S. Scott and H. Varian, "Predicting the Present with Bayesian Structural Time Series," *International Journal of Mathematical Modelling and Numerical Optimisation,* vol. 5, no. 1, pp. 4-23, 2014.

[9] G. E. P. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, Time Series Analysis Forecasting and Control Fifth Edition, Hoboken: John Wiley & Sons, Inc, 2016.

[10] R. Hyndman, G. Athanasopoulos, C. Bergmei, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos and S. Razbash, "Package ‘forecast’ Forecasting Functions for Time Series and Linear Model," CRAN-R, 2019.

[11] S. L. Scott, "Package ‘bsts’ Bayesian Structural Time Series," CRAN-R, Texas, 2019.

[12] I. Jaya, H. Folmer, B. N. Ruchjana, F. Kristiani and Y. Andriyana, "Modeling of Infectious Disease: A Core Research Topic for The Next Hundred Year," in *Regional Research Frontiers*, US, Springer, 2017, pp. 681-701.

[13] L. A. Gil-Alana, "Statistical Modeling of the Temperatures in the Northern Hemisphere Using Fractional Integration Techniques," *American Meteorological Society,* vol. 1, no. 1, pp. 5357-5369, 2005.