Business Analytics Report
MSc Business Analytics 2019-20
Imperial College Business School

Predictive Analytics in Weather Forecasting using Time Series Models: A Case Study for Aomori City in Japan

Konstantinos Paganopoulos
United Kingdom
August 2020

CID: 01769789
Word Count: 5054

Abstract

This paper presents a study of time series models for forecasting the weather of Aomori city in Japan. The report illustrates how time series can be used to build a reliable weather forecasting model. The objective is to explore regional weather patterns, examine the existence of climate change in Aomori during the 21st century and forecast mean monthly temperature for the next year with an average standard deviation error below 1°C. Eighteen-year data (2001-2018) of mean monthly temperature from the Japan Meteorological Agency are used. Auto Regressive Integrated Moving Average (ARIMA) and Holt-Winters Exponential Smoothing (ETS) models are developed. To produce more accurate results, data is split in train and test set. The best model is determined according to accuracy on the test set. According to RMSE (Root Mean Square Error), the ETS model performs more accurately. It predicts 2019 mean monthly temperature for Aomori with an average standard deviation of approximately 0.5 degrees Celsius. A steady increase in the mean annual temperature of Aomori during 2001-2018 indicates the presence of climate change during the 21st century.

1. Introduction

Weather forecasting systems belong to the most complex equations that computers attempt to solve. Data are collected mainly through a network of ground weather stations located around the globe. Weather forecasting is an issue of paramount importance not only for meteorology, scientific research and climate change detection, but also for the weather derivatives risk management and the agricultural production sector.

Regarding weather derivatives risk management, Sean D. Campbell and Francis X. Diebold state that (2003:p.4), 'weather forecasting is crucial to both the demand and supply sides of the weather derivatives market' by giving examples of firms exposed to weather risks. As for the agricultural production sector, Malgorzata Murat et. al claim that (2018:p.253), 'the prediction of the future courses of meteorological quantities on the basis of historical time series is important for agrophysical modelling (Lamorski et al.,

2013; Baranowski et al., 2015; Murat et al., 2016; Krzyszczak et al., 2017).' Malgorzata Murat et. al also mention that (2018:p.253), 'all the crop production models are highly sensitive to climatic and environmental variations (Fronzek et al., 2018; Pirttioja et al., 2015; Porter and Semenov, 2005; Ruiz-Ramos et al., 2018).' In terms of climate change, studies on climatic conditions have increased significantly in the last decades as the need to moderate the rise of the global mean temperature becomes imperative. Climate change can have a detrimental effect not only on earth but also on humans. The increasing temperatures are prone to affect crop seasons and food safety as well as the spread of diseases resulting in an increasing risk for people (UNFCCC, 2006).

In terms of past literature related to the topic, Holt (1957) implemented a linear trend method to allow forecasting data with trends. In 1960 Holt and Winters extended Holt's approach to capture seasonality. At a later stage, Cook and Wolfe (1991) used neural networks to predict mean temperature. Afterwards, Özelkan and Duckstein (1996), performed regression analysis and fuzzy logic when they examined the relationship between atmospheric circulation and precipitation. Liu and Chandrasekar (2000), used fuzzy logic and a neuro-fuzzy system to classify a hydrometeor type based on polarimetric radar measurements. Tektas (2010) forecasted weather parameters using Adaptive Neuro-Fuzzy Inference System (ANFIS) and Autoregressive Integrated Moving Average (ARIMA) models, with ANFIS yielding far better results. In a study similar to the present report, Garima and Bhawna (2017), predicted daily temperature using ARIMA and ETS models.

Therefore, forecasting temperature on a seasonal time scale has been attempted by numerous researchers by various techniques at different times all over the world. It is a difficult task to forecast mean monthly temperature that requires a very careful choice and a thorough implementation of the appropriate weather forecasting models.

In the first part of this study, descriptive statistics regarding weather in Aomori are illustrated via Python, R and Tableau software. Later, the relevant time-series theory is explained along with literature references. In the main part of the report, 2 models are implemented using R to predict mean monthly temperature in Aomori, and their findings

are finally presented in the conclusion. As previously stated, the aim of this report is to explore weather patterns and investigate if there is any climate change effect during the 21st century in Aomori city, so that it can later be used by meteorological agencies and various technocrats for accurate weather forecasts.

2. Case Study Area

Temperature is measured with higher accuracy compared to other weather parameters (Tektas, 2010). Hence, previous mean monthly temperature data concerning Aomori city have been used from the Aomori station of Japan Meteorological Agency. Regarding the time interval that is examined, data include a period of 18 years, from January 2001 to December 2018. The location of the area studied is shown in *Picture 1* (Aomori Prefectural Government, 2020).

2.1 Case Study Descriptive Statistics

When analysing the data of Aomori city weather station for 2001-2018, some interesting weather patterns and characteristics are observed. Visualising the correlation matrix between mean temperature, precipitation, mean wind speed, mean relative humidity and mean atmospheric pressure by month from 2001 to 2018 in a heatmap, shows that mean temperature is highly negatively correlated (-0.79) with mean wind speed (*Figure 1*). This is expected, since heat waves in general occur during periods of very low wind speed. Moreover, one can notice that mean temperature is also negatively correlated (-0.57) with mean pressure, indicating that probably winter months in Aomori are combined with higher atmospheric pressure, whilst summer ones with lower. The proof of the above is examined in detail in the following descriptive statistics.

More specifically, mean temperature in Aomori for the period 2001-2018 equals 10.7°C (*Table 1*). It forms a peak of 23.4°C during August and reaches its minimum of -1.1°C during January (*Figure 2*). As for the presence of climate change and its effect on the average annual temperature, a steady increase is observed from 2001 to 2018 (*Figure 3*).

Mean annual temperature increases from 9.9°C in 2001 to 11.0°C in 2018, with an earlier maximum of 11.5°C in 2015. Regarding the highest temperature in this period, that is 36.6°C in August 2010 (*Figure 4*) and the lowest is -10.9°C in January 2014 (*Figure 5*).

As for precipitation, it is observed (*Figure 6*) that from late winter until early summer and especially during spring, precipitation that falls in Aomori city is reduced significantly reaching its lowest in May, which is 67.3 millimeters (mm). Precipitation increases from July and forms an early peak on August with a mean value of 149.6 millimeters (mm) per month. Another peak is observed during December, which is the month of the year with the highest precipitation, with a value of 169.1 millimeters (mm) on average from 2001 to 2018. Notice that during winter the peak of the monthly precipitation is combined with a mean temperature that is close to 0°C, therefore heavy snowfalls occur during winter in Aomori. It is also worth mentioning that the minimum precipitation for a month during the examined period is 9 millimeters (mm), whilst the maximum is 345 millimeters (mm). Therefore, there is no month in the period from January 2001 until December 2018, with

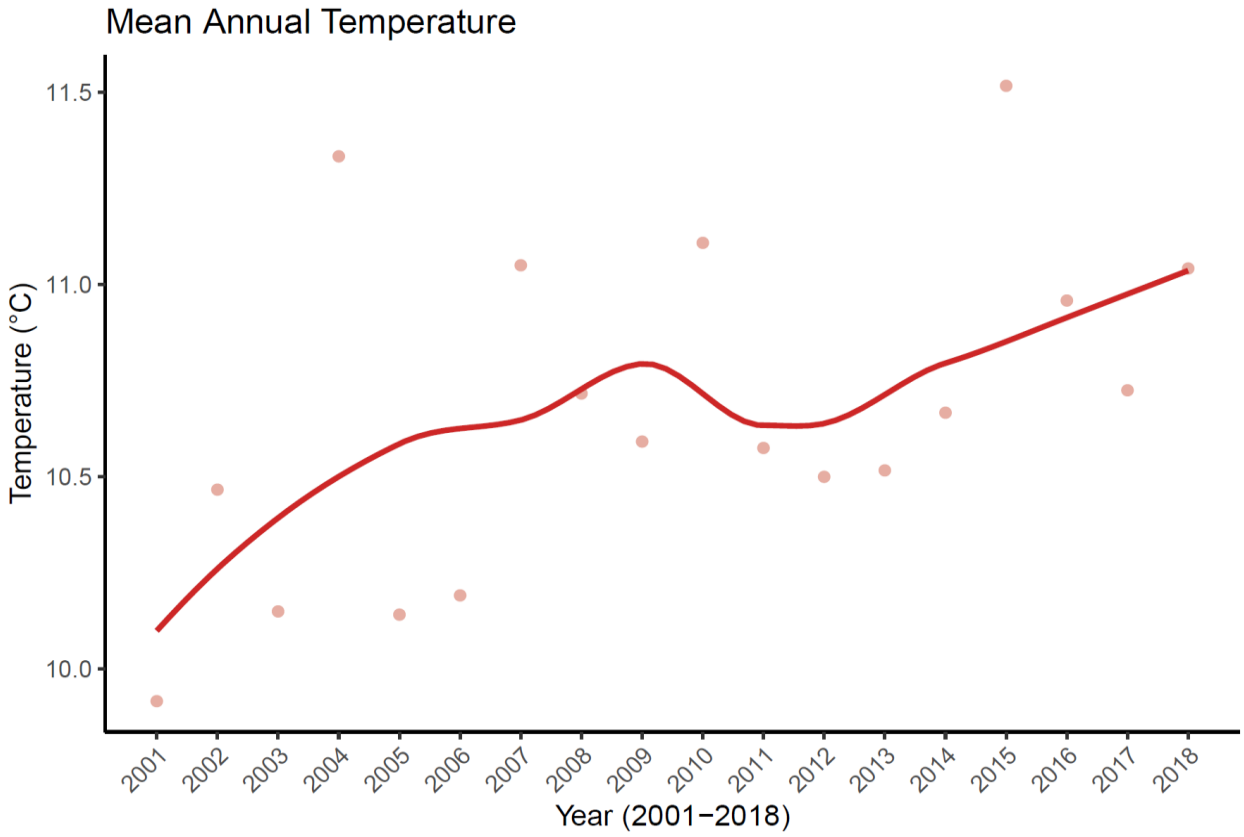

Figure 3: Mean temperature in Aomori city of Japan from 2001 to 2018 by year.

no precipitation in Aomori, with the maximum monthly precipitation of 345 millimeters (mm) consisting of 25% of mean annual precipitation (*Table 1*). Mean monthly precipitation is equal to 117.9 millimeters (mm), hence 1414.8 millimeters (mm), fall on average in Aomori city per year. As for the severity of the precipitation, the maximum highest daily values are mainly observed during autumn (mostly rainfalls), and the lowest during late winter or spring (*Figure 7*). This is justified by the fact that sea level temperature at "Mutsu Bay" near Aomori city is higher during Autumn resulting in more severe phenomena. On the other hand, in late winter and spring when sea level temperature is at its lowest, severe rainfalls are rare.

Concerning wind, it is clearly seen that Southwest (SW) and South-southwest (SSW) wind directions are the most dominant ones, summing up to a total of 74.35% out of all monthly dominant wind directions occurred at least once in 2001-2018 period (*Figure 8*). In terms of how wind direction changes per month, wind blows from the Southwest (SW) or South-southwest (SSW) directions with a higher frequency during all periods of the year, except for the three summer months. During the summer East-northeast (ENE) or North (N) wind directions are more frequent (*Figure 9*). Those East (E) wind directions are called "Yamase" and often result in cooler summers. As for whether dominant wind direction has changed over the years, in (*Figure 10*) a shift from Southwest (SW) to South-southwest (SSW), as the wind with the highest frequency, is observed during the last years.

Examining wind speed per wind direction (*Figure 11*), West-northwest (WNW), West (W), and West-southwest (WSW) winds, have the highest speed. This is expected, since those wind directions occur when a cold front pass above an area, and high winds are frequently observed during cold fronts of very deep low barometric systems. When correlating wind with temperature (*Figure 12*), it is seen that North-northeast (NNE) winds are associated with the highest average temperature (23.4°C) and West-northwest (WNW) winds with the lowest (-2.5°C). This is explained as North-northeast (NNE) winds are more frequent during summer months, whereas West-northwest (WNW) occur mostly after the passage of a cold front, when temperature falls significantly. To conclude, observing *Table 1* reveals that mean wind speed in Aomori is equal to 13.3

kilometers per hour (km/h), with a mean highest wind gust per month of 80.6 kilometers per hour (km/h), which equals 9 in Beaufort Scale. The highest wind gust for the examined period is 130.7 kilometers per hour (km/h), or 12 Beaufort. Therefore, Aomori city of Japan faces with very strong winds throughout the year.

Relative humidity is quite steady, however an increase during summer months and a decrease during spring is observed (*Figure 13*). Mean humidity levels fall slightly below 65% during April and slightly exceed 80% during July. The minimum is explained since the lowest amount of precipitation falls during spring, and the summer maximum, since Aomori city is affected by "Oyashio" current during summer, which occasionally results in the creation of fog. The mean monthly humidity level is approximately equal to 75% (*Table 1*), which shows that Aomori has a humid climate.

Mean atmospheric pressure is equal to 1012.8hPa, with a highest value of 1021.3hPa and a lowest of 1005.8hPa in the recorded period. Hence, pressure is quite steady, with not extreme fluctuations. Higher barometric pressure values are observed from late autumn until late winter, reaching a peak of 1016.5hPa during November, and lower during summer, reaching a minimum of 1008.0hPa during July (*Figure 14*). The higher values during autumn and winter in Aomori are caused by the relatively cold and wet conditions experienced in this region compared to summer. Since cold air is denser than warm air, it tends to sink, and thus has a higher pressure than warm air.

In conclusion, Aomori has in general a humid climate with warm summers and cold, though not extreme, winters.

3. Time Series Decomposition

According to Athanasopoulos et al. (2018:p.197), 'it is often helpful to split a time series into several components, each representing an underlying pattern category.' Time series consist of three components: a trend-cycle component, a seasonal component, and a remainder component (containing anything else included in the time series). More specifically, seasonal patterns (S_t) are cyclical fluctuations related to calendar (e.g. time of the year or day of the week) or business cycle, that are always of a fixed and already

known frequency. A trend (T_t) exists when there is a long-term movement (increase or decrease) in the mean of the data. The remainder component (R_t) or microscopic part accounts for other random or unsystematic fluctuations represented by residuals. Note that trend and cycle are combined into a single trend-cycle component called the trend for simplicity reasons. The time series that contain the above components are usually described as $y_t = S_t + T_t + R_t$, where y_t is the data, S_t is the seasonal component, T_t is the trend-cycle component, and R_t is the remainder component, at period t.

3.1 Autoregressive Models

In autoregression models, the value of the variable of interest is forecasted via a linear combination of the previous values of that variable. Hence, an autoregressive model of order p is written as $y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t$, where ε_t is white noise. Notice that this is like a multiple regression but with lagged values of y_t as predictors. Reference to autoregressive models of order p is denoted as AR(p).

3.2 Moving Average Models

Moving average models use previous forecast errors instead of previous values of the forecast variable in a regression. Hence, a moving average model is written as the following: $y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$, where ε_t is white noise. Note that each value y_t of predictors is as a weighted moving average of the previous forecast errors. Reference to moving average models of order q is denoted as MA(q).

3.3 Non-seasonal Arima Models

Combining differencing with autoregression and moving average models, results in non-seasonal AutoRegressive Integrated Moving Average (ARIMA) models. Arima models aim to describe the autocorrelations in the data. An ARIMA model is written as follows: $y_t' = c + \varphi_t y_{t-1}' + \ldots + \varphi_p y_{t-p}' + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$, where y_t' is the differentiated time series. Note that each value of the predictors includes not only lagged values of y_t but

also lagged errors. Reference to that models is denoted as ARIMA(p, d, q), where p is the order of the autoregressive part, d is the degree of first differencing involved and q is the order of the moving average part.

3.4 Exponential Smoothing Models

The idea behind the exponential Soothing is that the forecasts by this type of models are weighted averages of past observations and the weights decay exponentially with time. Therefore, the more recent the observations, the larger the weights compared to the earlier ones, resulting in smarter and thus better forecasts (Garima et al., 2016).

3.4.1 Holt's Linear Trend Models

Holt (1957) extended simple exponential smoothing, according to which forecasts for the future are equal to the last observed or average value of the time series, to be able to forecast those with trends (Athanasopoulos et al., 2018). This type of models consist of a forecast equation and two smoothing equations (one level and one trend equation): Forecast equation: $y_{t+h|t} = \ell_t + hb_t$, Level equation: $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$, Trend equation: $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$, where ℓ_t is an estimate of the level of the data during time t, b_t is an estimate of the trend of the data during time t, α is the smoothing parameter for level, $0 \leq \alpha \leq 1$, β^* is the smoothing parameter for trend, $0 \leq \beta^* \leq 1$. Note that ℓ_t is a weighted average of the observation y_t and the one-step-ahead training set forecast for time t, or $\ell_{t-1} + b_{t-1}$. The trend equation reveals that b_t is a weighted average of the estimated trend during time t based on $\ell_t - \ell_{t-1}$ and b_{t-1}, the past estimate of that trend (Athanasopoulos et al., 2018).

3.4.2 Holt-Winters' Seasonal Additive Models

As Athanasopoulos et al. (2018:p.257) state, 'Holt (1957) and Winters (1960) extended Holt's approach to capture seasonality'. Holt-Winters seasonal method consists of the forecast equation and three smoothing equations — one for level ℓ_t, one for trend b_t, and

one for seasonal component s_t, with corresponding smoothing parameters α, β^* and γ. The component form for the additive method is the following: $y_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)}$, $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$, $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{\tau-1}$, $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$, where k is the integer part of $(h-1)/m$, which guarantees that the estimates of the seasonal indices used for forecasting derive from the final year, and m stands for the frequency of seasonality (Athanasopoulos et al., 2018). Note that the level equation is a weighted average of the seasonally adjusted observation $(y_t - s_{t-m})$ and the non-seasonal forecast $(\ell_{t-1} + b_{t-1})$ for time t. The trend equation is like that in Holt's linear models. The seasonal equation is a weighted average of the current seasonal index, $(y_t - \ell_{t-1} - b_{t-1})$, and the seasonal index of the same season in the previous year.

4. Case Study Arima Modelling

The initial data set consists of monthly weather data from 1st January 2001 until 31st December 2019 (228 observations). Each observation includes fourteen values: year, month, mean temperature, mean maximum temperature, mean minimum temperature, maximum temperature, minimum temperature, precipitation, highest daily precipitation, mean wind speed, highest wind gust, dominant wind direction, mean relative humidity and mean atmospheric pressure in that particular month. The monthly temperature data from January 1st, 2001 to December 31st, 2018 is converted into monthly time series data (216 observations). Frequency shows the number of observations per unit of time. As the objective is to forecast mean temperature by month for the next year, time series data are split into training set (years 2001-2017), consisting of 204 observations (89%) and validation set (year 2018), consisting of 12 observations (5.5%). At the end of the analysis it is checked how the final model is performing in the test set (year 2019), consisting of the last 12 observations (5.5%).

Regarding the implementation of the Arima model in the examined case study area, first, in order to run Arima model in the training set, it is examined if the data are stationary or not. Observing the pattern of the time-series data in a graph is very useful in order to understand if there is stationarity in the data. Plotting the time series data for 2001-2017

clearly shows that the series are not stationary (*Figure 15, 16*), as cyclical fluctuations related to calendar are observed. These fluctuations have a frequency of one year, which indicates a seasonal pattern. In order to get rid of seasonality first-order difference is needed. The need for differentiating for seasonality one time is proved when running the ndiffs() and nsdiffs() functions of R software.

Performing Augmented Dickey-Fuller (ADF), Phillips-Perron Unit Root (PP) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests confirms that data are not stationary. More specifically, test statistic of KPSS test is equal to 0.0497, which is less than the critical values at 1%, 5% and 10% significance level. Therefore, the null hypothesis that series are stationary is not rejected (*Table* 2) and the temperature series are stationary. Running the ndiffs() and nsdiffs() functions after differentiation, shows that no first-order differences are needed anymore. Note that by default ndiffs() is based on KPSS test. In differentiation, the argument lag = 12 is used to capture annual seasonality.

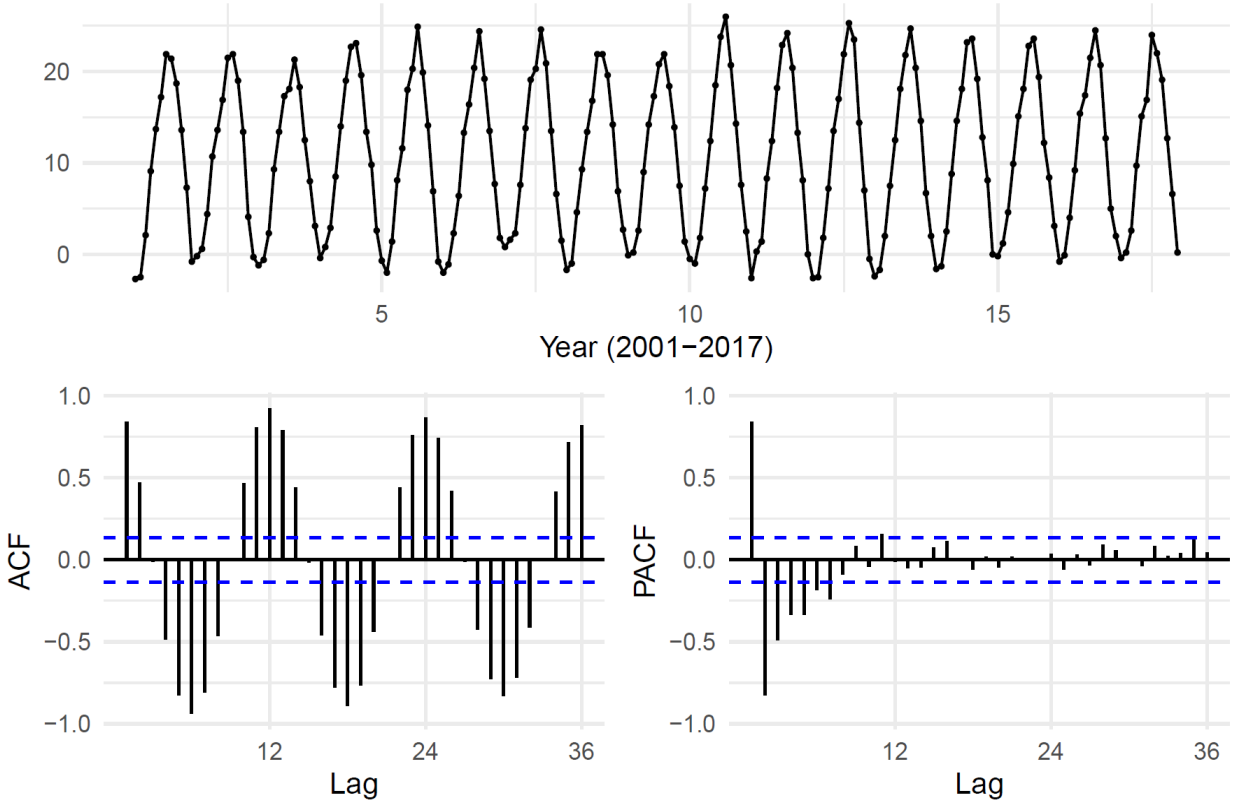

Figure 16: Mean monthly temperature time series data from 2001 to 2017 by year and Autocorrelation and Partial Autocorrelation function for white noise series in Aomori city.

After differentiation, visual observation of whether time series are stationery follows. The series look stationary (*Figures 17, 18*) and the next step is to determine the optimal orders of MA and AR components by plotting the ACF and PACF of the time series (*Figures 19, 20*). However, the same results are obtained by running the auto.arima function of R. As Athanasopoulos et al. (2018:p.321) state, 'the default procedure of auto.arima function uses some approximations to make the search faster.' The above-mentioned approximations are avoided with the command approximation = FALSE. More models will be examined if the command stepwise = FALSE is used. Note that it is possible that the minimum AIC (best) model will not be found either because of these approximations, or because of the stepwise procedure. The argument d in the auto.arima function denotes the number of times first order differencing is required, and that of D the number of times seasonal differencing is needed. Since, seasonal differencing but no first differencing was applied, then the arguments d = 0 and D = 1 are used. Based on AIC score, the following models are considered as possible models to represent the original series: ARIMA(1,0,0)(0,1,1) with AIC = 622.9, ARIMA(0,0,2)(0,1,1) with AIC = 623.3 and ARIMA(2,0,0)(0,1,1) with AIC = 624.4, all without drift. The analysis continues by choosing the best of the three possible models, in other words the one with the lowest AIC score, namely ARIMA(1,0,0)(0,1,1) or ARIMA(1,0,0)(0,1,1)[12] without drift and an AIC of 622.9, as the candidate model.

The in-sample performance or fit of the model is evaluated with accuracy() function of R, which summarizes various measures of fitting errors (Paganopoulos, 2020). Furthermore, in the post-estimation analysis, residual plots, time series and ACFs are examined to confirm that there is no warning signal. In this case, residuals have a zero mean, constant variance, and are distributed symmetrically around mean zero (*Figure 21*). As a result, the forecasting section can now become the next step of the procedure followed. Continuing with the forecasting part of the candidate model, the in sample forecast for 2018 predicted by forecast() function of R is seen in *Figure 22*.

In order to test how the model performs for validation set earlier observations are used for training, and more recent observations are used for validation. The first 204

observations of data are used for training and the last 12 for validation and the model with the lowest AIC from auto.arima function is chosen to be the final model.

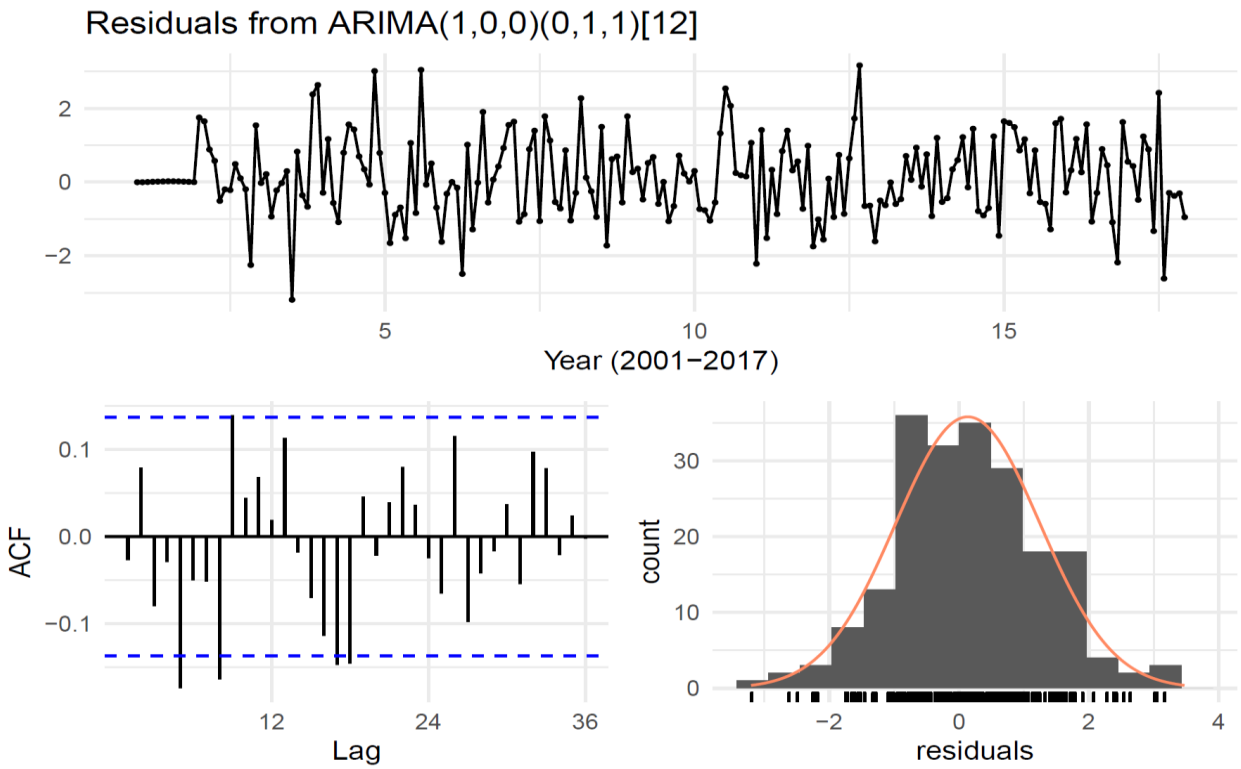

Figure 21: Residuals from Arima(1,0,0)(0,1,1) model and Autocorrelation function.

Regarding the accuracy of the best model, it performs with a RMSE of 0.89 in the validation set, which would be considered as being within acceptable range (*Table 3*).

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0.13	1.12	0.87	-∞	+∞	0.71	-0.02	NA
Validation set	0.36	0.89	0.71	5.77	18.33	0.58	-0.24	0.45

Table 3: Training and validation set accuracy of Arima(1,0,0)(0,1,1) model.

The best model is then trained to the whole data set and the mean monthly temperature for the next year in Aomori city is forecasted by forecast() function of R (*Table 4*).

5. Case study Holt-Winters Modelling

In every model implementation for time series analysis, the first step is to visually inspect the time series. As Wu (2020) claims, 'in this regard the stl() function is quite useful. It

decomposes the original time series into trend, seasonal factors, and random error terms.' It is worth pointing out that the importance of the various components is provided by the grey bars on the right side of the four plots (*Figure 23*).

Note that the grey bar of the trend component is significantly larger than that of the original time series data component. What this shows is that trend contributes to the variation of the original data marginally. On the other hand, the bar of the seasonal panel is very small, even smaller than that of the random error term, implying that the contribution of seasonality is huge. In other words, seasonality in the data is high.

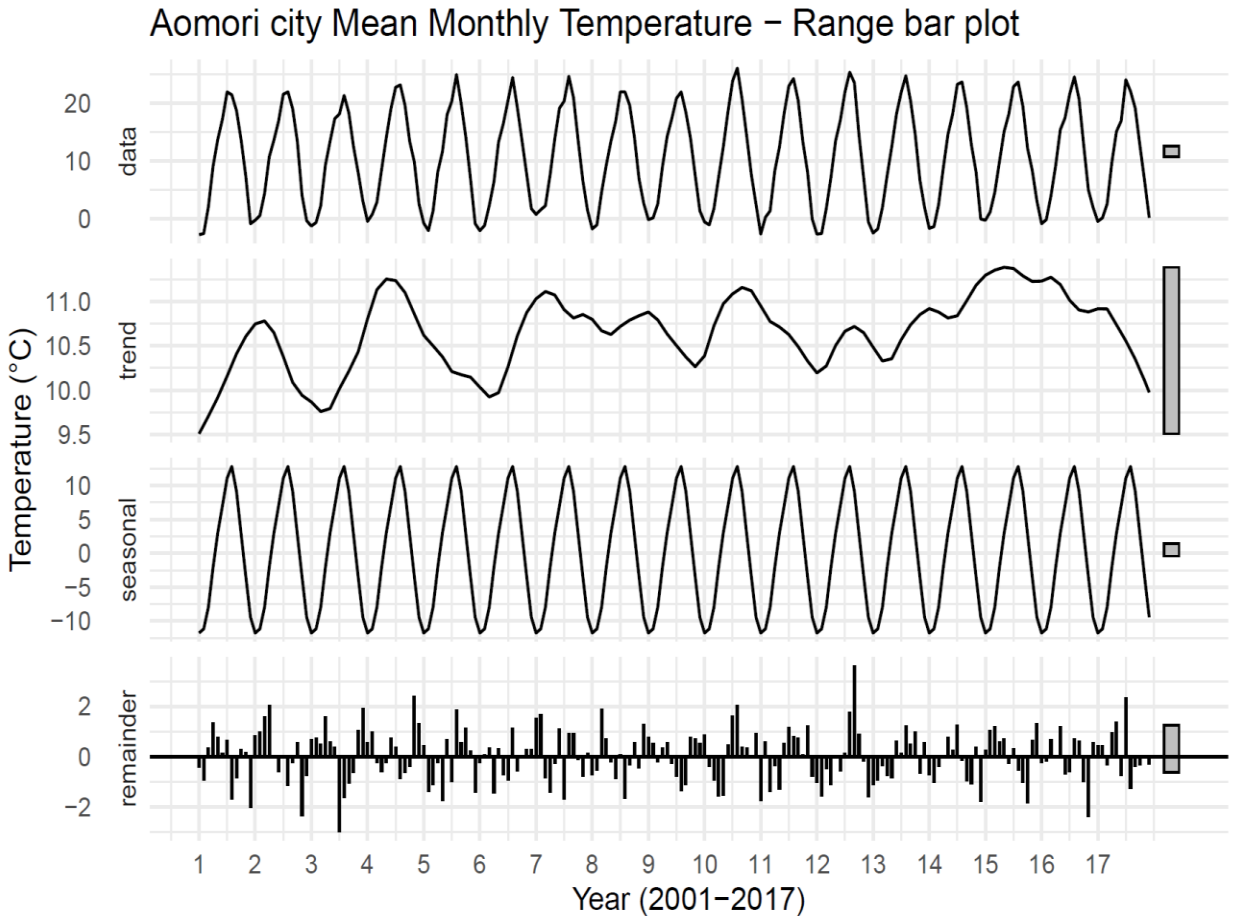

Figure 23: Range bar plot; mean monthly temperature time series data, trend, seasonal and remainder component from 2001 to 2017 by year in Aomori city of Japan.

In the ets() function of R, parameters are estimated by maximizing the likelihood function (Wu, 2020). In order to choose the model in the ets() function, three letters need to be specified. The trend and seasonality panel in *Figure 23* shows that seasonality is high, but

trend is minor. Running the ets function with the argument model = "ZZZ" results in finding the best model. According to the results, the best model is the Additive Holt-Winters' method or ETS(A,A,A).

Using accuracy() function shows the in-sample performance of the best model and forecast() function of R predicts the future values. Note that the AIC score is used again to determine the best model in terms of in-sample performance. Regarding the accuracy of the best model, it performs with a RMSE of 0.85 in the validation set, which would be considered again as being within acceptable range (*Table 5*). Note that again RMSE is chosen as an accuracy metric, as it is useful when large errors are undesirable.

The best model, ETS(A,A,A), is trained in the whole data set and the mean monthly temperature for the next year in Aomori city is calculated (*Table 6*).

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-0.01	1.14	0.92	-∞	+∞	0.75	0.34	NA
Validation set	0.01	0.85	0.66	10.49	17.43	0.54	-0.27	0.57

Table 5: Training and validation set accuracy of ETS(A,A,A) model.

6. Case Study Model Comparison

It is believed that ARIMA models are more general than the exponential smoothing ones. Linear exponential smoothing models form a subcategory of ARIMA models; however, the non-linear exponential smoothing models have no ARIMA equivalents. Of course, there are also many ARIMA models that have no exponential smoothing equivalents. Notice that all ETS models are non-stationary, whereas some ARIMA models are stationary and some non-stationary. Moreover, the ETS models that have seasonality need two times of differencing to be transformed into stationary, whilst the rest ETS models need only to be differentiated once for them to be stationary. Comparing the AIC score is useful for selecting between models. On the other hand, it is not recommended to compare an ETS and an ARIMA model by AIC score, because they belong in different model classes, and likelihood is not calculated in the same way (Athanasopoulos et al., 2018).

In terms of comparison of the best two candidate models for the examined case study, the ARIMA(1,0,0)(0,0,1) with the ETS(A,A,A), the Holt-Winters model yields a lower RMSE score in the test set than the Arima model, thus a better accuracy. More specifically, ETS has a RMSE of 0.85, whereas Arima a RMSE score of 0.89 (*Tables 3, 5)*. Notice that again RMSE is chosen as an accuracy metric for the final model, since large errors are undesirable. Plotting the time series for train and validation set and the forecasts from the two candidate models shows each model's monthly temperature prediction in Aomori City for 2019 or alternatively the fit of the two models to the data (*Figures 24, 25*).

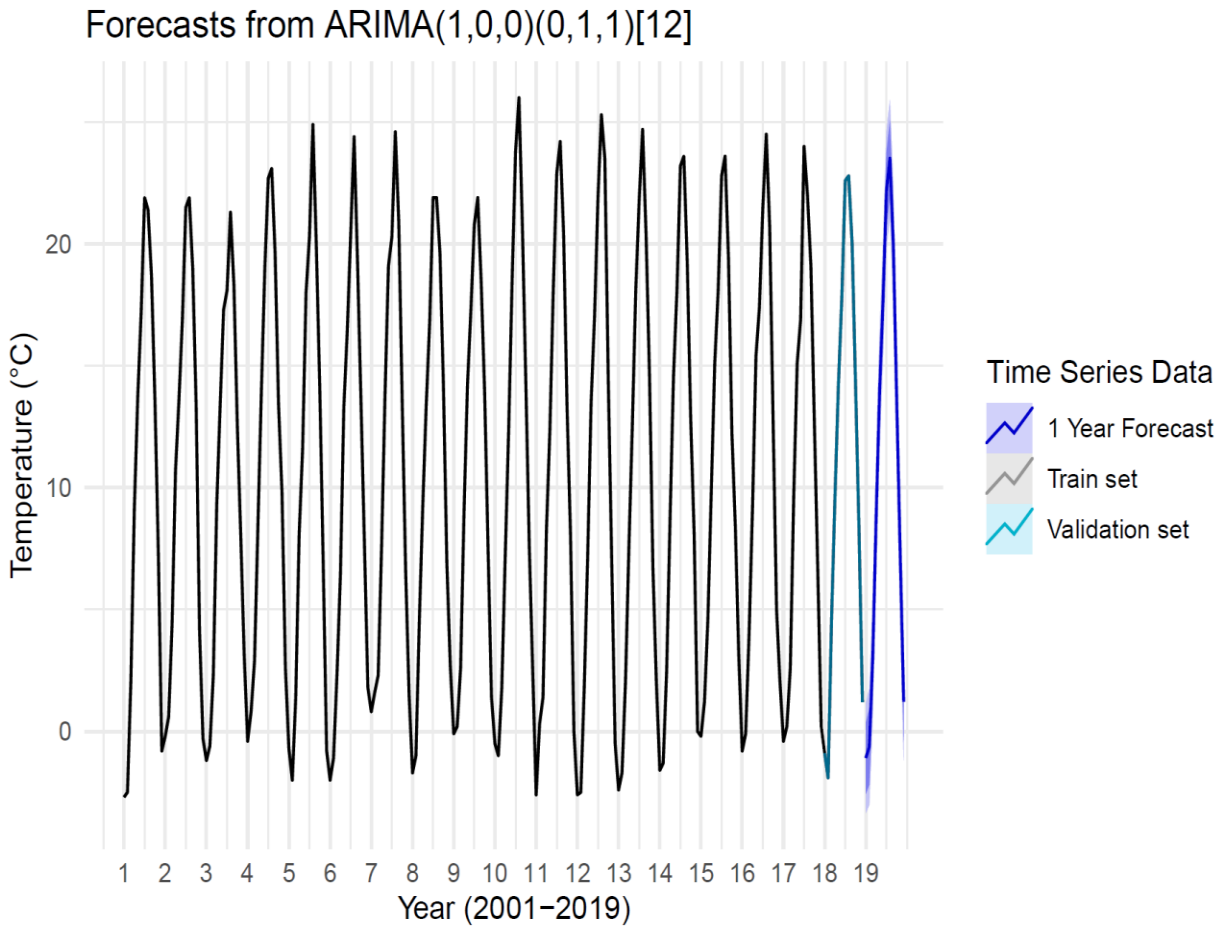

Figure 24: ARIMA(1,0,0)(0,1,1) forecast for 2019 mean monthly temperature in Aomori city.

As previously mentioned, since ETS(A,A,A) has a better (lower) RMSE (0.85 vs 0.89) respectively, choosing the ETS(A,A,A) as the final weather forecasting model seems the best choice. Comparing the values predicted from the two models, with the actual mean monthly temperatures for 2019 reveals how accurate the forecasts are. In order to

compare the ARIMA and the ETS results with the actual data, the standard deviation or the absolute difference between the actual mean monthly temperatures in 2019 and the predictions is calculated. The absolute difference or standard deviation is named as error (arima and ets error respectively) for simplifying reasons (*Table 7*).

Clearly as seen in *Table 7,* Holt-Winters model has not only a lower RMSE score, but also predicts mean monthly temperature with a lower absolute difference or standard deviation from the actual values compared with the Arima one. In any case, the absolute distance from actual values is the most valuable characteristic in a weather forecasting model, hence the ETS model outperforms the Arima one.

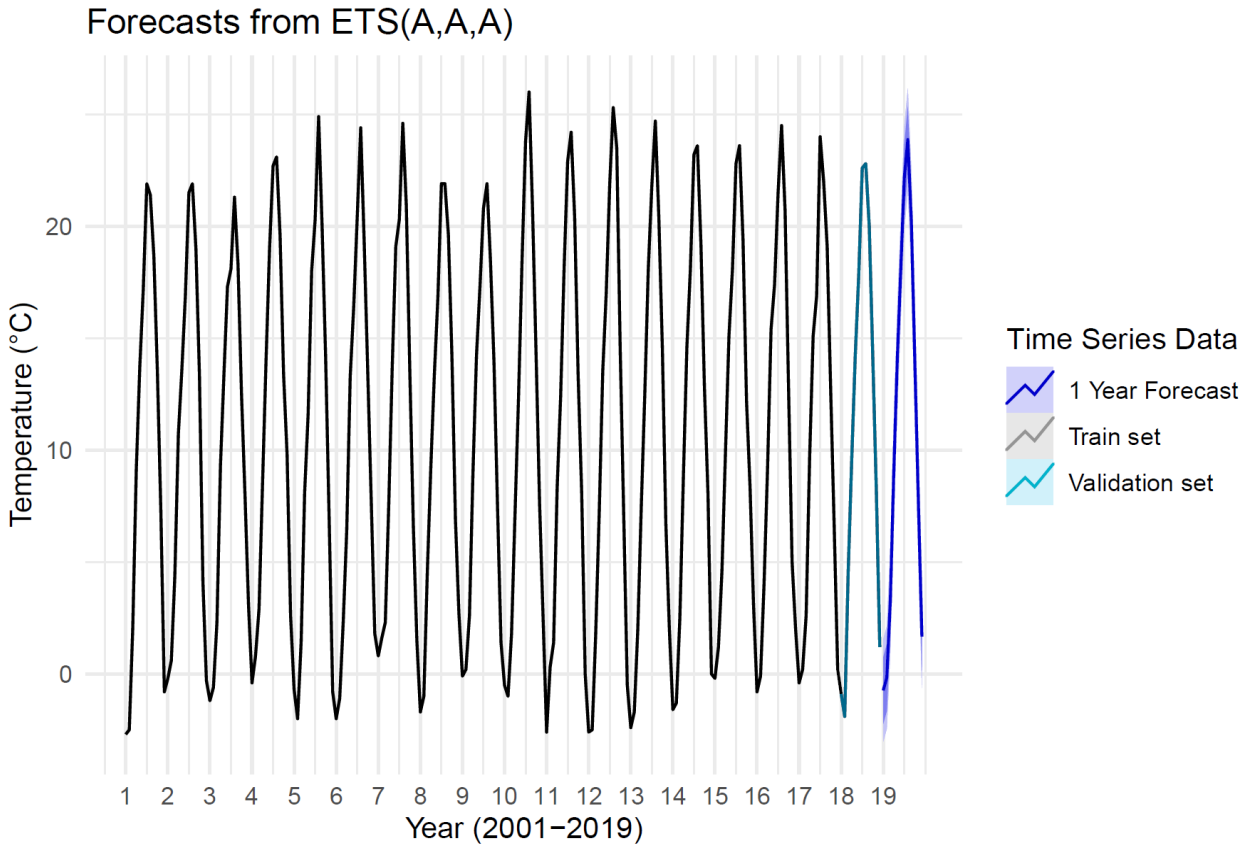

Figure 25: ETS(A,A,A) forecast for 2019 mean monthly temperature in Aomori city.

Calculating the mean standard deviation of the forecasted values from the actual ones, results in a mean error of 0.87 regarding ARIMA model, and a mean error of only 0.67 regarding ETS model. Thus, the Holt-Winters model performs quite better on the test set. Note that the approximately 0.5 degrees Celsius difference of the forecast from the actual

values denotes a very accurate result. In particular, the ETS model predicted that January will be the coldest month with -0.7°C as a mean temperature, which is almost identical to the actual mean temperature of -0.6°C for January 2019. Regarding the hottest month, the ETS forecast of 23.9°C for August is only 0.9°C below the actual reported mean of 24.8 degrees Celsius. It is also worth mentioning that the ETS model predicted mean temperature for June 2019 with 100% accuracy, and that the absolute difference between the forecast and the actual mean temperature, is in nine of the twelve months lower or equal than 1 degree Celsius. On the other hand, the ARIMA model even though performs a little worse, it is still a very accurate and thus a reliable model for forecasting mean monthly temperature, as it shows approximately the same monthly absolute difference errors with the ETS model, especially during autumn and spring months. Notice that during late summer and early autumn months, both models predict mean monthly temperature with a slightly larger error, and the reason for that is probably the higher climate change effect during the aforementioned months. This climate change effect potentially leads to temperature values higher than the predicted ones.

Month (2019)	Arima Forecast	Ets Forecast	Actual Data	Arima Error	Ets Error
January	-1.1	-0.7	-0.6	0.5	0.1
February	-0.6	-0.2	0.0	0.6	0.2
March	3.0	3.2	3.8	0.8	0.6
April	8.8	9.0	8.5	0.3	0.5
May	14.0	14.1	15.8	1.8	1.7
June	17.7	18.1	18.1	0.4	0.0
July	22.2	22.2	22.1	0.1	0.1
August	23.5	23.9	24.8	1.3	0.9
September	20.0	20.2	21.2	1.2	1.0
October	13.4	13.9	15.1	1.7	1.2
November	7.2	7.7	6.3	0.9	1.4
December	1.2	1.7	2.0	0.8	0.3

Table 7: Final Forecast of ARIMA(1,0,0)(0,0,1) and ETS(A,A,A) model and actual data for 2019 mean monthly temperature of Aomori city in Japan, with respective absolute difference from actual data as error.

The final forecast from the two models along with the actual mean monthly temperature for 2019 in Aomori city of Japan are depicted in *Figure 26*.

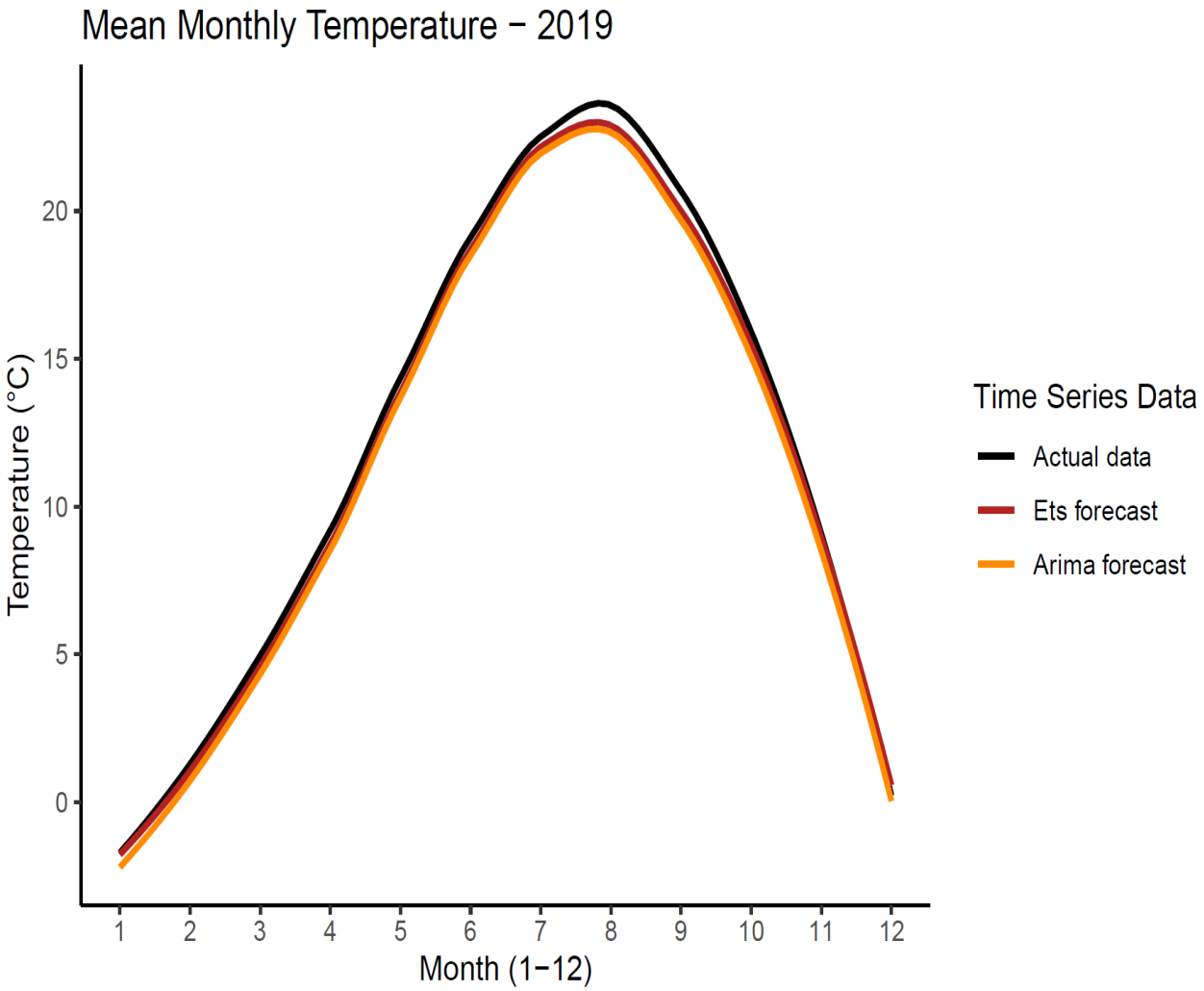

Figure 26: Mean Monthly Temperature (2019) for Aomori city in Japan; Comparison of actual data with predicted values from Ets and Arima models.

7. Limitations

In this study data from Aomori city weather station for the period between 2001 and 2018 are used. One can speculate that eighteen years of data might not be enough to draw insights and conclusions about Aomori. In meteorology thirty years of data are mainly used to depict the past average of temperature in "ensembles" (Wetterzentrale.de 2020), however eighteen years are enough to show a clear pattern regarding mean temperature

in Aomori. As previously stated, temperature is measured with a higher degree of accuracy among all weather parameters and this is the main reason why it is chosen. Further examination of climate change effect in other weather parameters besides temperature can be examined in future research. Moreover, descriptive statistics in this paper reveal that temperature is the weather parameter that shows the largest difference during the examined period in Aomori, and thus is the one investigated in more detail. Furthermore, a similar analysis can be used to forecast mean monthly temperature of 2020 or even that of future years. In this analysis, in order to produce more accurate results, predicting the mean monthly temperature only for the next year is decided. Nevertheless, in meteorology and weather forecasting, the ability to forecast weather drops significantly as the future time interval is increasing. The so-called "ensembles" attempt to forecast long-term weather parameters, but one should be aware that the predictability is much worse as time passes (Tarek, 2016). Consequently, the need for long-term forecasts can potentially lead to the implementation of inaccurate and thus unreliable weather forecasting models, which ought to be considered by every future researcher.

8. Conclusions and Recommendations

This study reveals that mean annual temperature has risen by 1.1°C from 2001 to 2018 in Aomori city of Japan, indicating climate change effect. Regarding regional weather patterns, mean temperature reaches a peak of 23.4°C during summer and falls at a minimum of -1.1°C in winter. Precipitation is uniformly distributed throughout the year, except for a period between late winter and early summer when is reduced. Severe rainfalls occur during autumn due to relatively high sea level temperature at "Mutsu Bay". Winds blow mainly from Southwest directions, however a slight shift to South-southwest directions occurs in the last years. During summer, winds blow from East-northeast or North directions. Coldest and strongest winds are those from the West-northwest, as they appear immediately after cold fronts. Humidity increases during months with higher precipitation and decreases during those with lower; a slight increase

is observed in summer, due to fog incidents caused by "Oyashio" current. Pressure is lower during hot months and higher during cold ones, as cold air is denser than warm. Regarding predictive analytics, Holt-Winters Exponential Smoothing method with ETS(A,A,A) model is the most appropriate for forecasting mean monthly temperature in Aomori, showing an average standard deviation of 0.67°C from actual values. Japan Meteorological Service can use these findings not only to produce accurate forecasts of surface temperature but also to help technocrats of the derivatives risk management and the agricultural production sector. Similar analysis can be done in future research to forecast either mean monthly temperature or other weather parameters for a longer period.

9. Acknowledgements

We thank the Japan Meteorological Agency for providing the necessary data needed for this study. The contribution from ENDO Hiroya in the Office of International Affairs of Japan Meteorological Agency proved to be very valuable for this research, not only through providing access to extended meteorological data but also through the translation of multiple Japanese files to English.

References

Aomori Prefectural Government (2020) *Aomori Prefecture, Tohoku Area of Honshu Island, Japan.* Available from: http://www.sushitrainer.net/JapaneseInfo/aomori.htm [Accessed: 1st July 2020].

Baranowski P., Krzyszczak J., Sławiński C., Hoffmann H., Kozyra J., Nieróbca A., Siwek K., and Gluza A. (2015) Multifractal Analysis of Meteorological Time Series to Assess Climate Impacts. *Climate Research.* 65, 39-52. Available from: doi:10.3354/cr01321

Campbell, S. and Diebold, F. (2005) Weather Forecasting For Weather Derivatives. *Journal of the American Statistical Association.* 100, 469. Available from: doi:10.1198/016214504000001051

Cook, D. and Wolfe, M. (1991) A back-propagation neural network to predict average air temperatures, *AI Applications*. 5, 40-46. Available from: https://agris.fao.org/agris-search/search.do?recordID=US9125059 [Accessed: 5th July 2020].

Fronzek S., Pirttioja N., Carter T.R., Bindi M., Hoffmann H., Palosuo T., Ruiz-Ramos M., Tao F., Trnka M., Acutis M., Asseng S., Baranowski P., Basso B., Bodin P., Buis S., Cammarano D., Deligios P., Destain M.-F., Dumont B., Ewert F., Ferrise R., François L., Gaiser T., Hlavinka P., Jacquemin I., Kersebaum K.C., Kollas C., Krzyszczak J., Lorite I.J., Minet J., Minguez M.I., Montesino M., Moriondo M., Müller C., Nendel C., Öztürk I., Perego A., Rodríguez A., Ruane A.C., Ruget F., Sanna M., Semenov M.A., Sławiński C., Stratonovitch P., Supit I., Waha K., Wang E., Wu L., Zhao Z., and Rötter R.P. (2018) Classifying multi-model wheat yield impact response surfaces showing sensitivity to temperature and precipitation change. *Agricultural Systems*. 159, 209-224, Available from: doi:10.1016/j.agsy.2017.08.004

Garima, J. and Bhawna, M. (2017) A Study of Time Series Models ARIMA and ETS. *International Journal of Modern Education and Computer Science*. 4, 57-63. Available from: doi:10.5815/ijmecs.2017.04.07

Holt, C. E. (1957) Forecasting seasonals and trends by exponentially weighted averages. *Carnegie Institute of Technology, Pittsburgh USA*. 52. Available from: doi:10.1016/j.ijforecast.2003.09.015

Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. Available from: OTexts.com/fpp2. [Accessed 11th July 2020].

Lamorski K., Pastuszka T., Krzyszczak J., Sławiński C., and Witkowska-Walczak B. (2013). Soil water dynamic modeling using the physical and support vector machine methods. Vadose Zone Journal. 12 (4), Available from: doi:10.2136/vzj2013.05.0085

Liu, H. and Chandrasekar, V. (2000) Classification of Hydrometeor Type Based on Multiparameter Radar Measurements: Development of a Fuzzy Logic and Neuro Fuzzy

Systems and In-situ Verification, *J. Atmos. Ocean Tech.*, 17 (2), 140-164. Available from: doi:10.1175/1520-0426(2000)017<0140:COHBOP>2.0.CO;2

Krzyszczak J., Baranowski P., Hoffmann H., Zubik M., and Sławiński C. (2017) Analysis of Climate Dynamics Across a European Transect Using a Multifractal Method. *Springer International Publishing*. 103-116. Available from: doi:10.1007/978-3-319-55789-2_8

Murat M., Malinowska I., Gos M., and Krzyszczak J. (2018) Forecasting daily meteorological time series using ARIMA and regression models. *International Agrophysics.* 32, 253-264. Available from: doi:10.1515/intag-2017-0007

Murat M, Malinowska I., Hoffman H., and Baranowski P. (2016) Statistical modelling of agrometeorological time series by exponential smoothing. *International Agrophysics*. 30 (1), 57-66. Available from: doi:10.1515/intag-2015-0076

Ozelkan, E., Ni, F. and Duckstein, L. (1996): Relationship between monthly atmospheric circulation patterns and precipitation: Fuzzy logic and regression approaches. *Water Resources. Research.* 32, 2097–2103. Available from: doi:10.1029/96WR00289

Paganopoulos, K. (2020) Demand Forecasting for a Fast-Food Restaurant Chain*, Imperial College Business School.* Available from: https://github.com/kpaganopoulos/timeseries-forecasting [Accessed: 12th July 2020].

Pirttioja N., Carter T.R., Fronzek S., Bindi M., Hoffmann H., Palosuo T., Ruiz-Ramos M., Tao F., Trnka M., Acutis M., Asseng S., Baranowski P., Basso B., Bodin P., Buis S., Cammarano D., Deligios P., Destain M.-F., Dumont B., Ewert F., Ferrise R., François L., Gaiser T., Hlavinka P., Jacquemin I., Kersebaum K.C., Kollas C., Krzyszczak J., Lorite I.J., Minet J., Minguez M.I., Montesino M., Moriondo M., Müller C., Nendel C., Öztürk I., Perego A., Rodríguez A., Ruane A.C., Ruget F., Sanna M., Semenov M.A., Sławiński C., Stratonovitch P., Supit I., Waha K., Wang E., Wu L., Zhao Z., and Rötter R.P. (2015) Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. *Climate Research*. 65, 87-105, Available from: doi:10.3354/cr01322

Porter J.R. and Semenov M.A. (2005) Crop responses to climatic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 360 (1463), 2021-2035. Available from: doi:10.1098/rstb.2005.1752

Ruiz-Ramos M., Ferrise R., Rodríguez A., Lorite I.J., Bindi M., Carter T.R., Fronzek S., Palosuo T., Pirttioja N., Baranowski P., Buis S., Cammarano D., Chen Y., Dumont B., Ewert F., Gaiser T., Hlavinka P., Hoffmann H., Höhn J.G., Jurecka F., Kersebaum K.C., Krzyszczak J., Lana M., Mechiche-Alami A., Minet J., Montesino M., Nendel C., Porter J.R., Ruget F., Semenov M.A., Steinmetz Z., Stratonovitch P., Supit I., Tao F., Trnka M., de Wit A., and Rötter R.P. (2018). Adaptation response surfaces for managing wheat under perturbed climate and CO2 in a Mediterranean environment. *Agricultural Systems*. 159, 260-274, Available from: doi:10.1016/j.agsy.2017.01.009

Tarek S. (2016) Ensembles for the predictability of Average Temperatures. *University of Pittsburgh.* Available from: http://d-scholarship.pitt.edu/30320/ [Accessed: 9th August 2020].

Tektas, M. (2010) Weather Forecasting Using ANFIS and ARIMA MODELS. A Case Study for Istanbul. *Marmara University, Vocational School of Technical Sciences, Turkey*. 1 (51), 5-10. Available from: http://erem.ktu.lt/index.php/erem/article/view/58 [Accessed 7th July 2020].

United Nations Framework Convention on Climate Change (2006) Climate change: Impacts, vulnerabilities and adaptation in developing countries. Available from: https://unfccc.int/resource/docs/publications/impacts.pdf [Accessed: 4th July 2020].

Wetterzentrale.de (2020) *850 hPa Temp. in °C, 6h-Niederschlag in mm, Daten: Ensembles des GFS von NCEP*. Available from: http://old.wetterzentrale.de/pics/MT8_Athen_ens.png [Accessed: 6th August 2020].

Winters, P. R. (1960) Forecasting sales by exponentially weighted moving averages. *Management Science.* 6, 324–342. Available from: doi:10.1287/mnsc.6.3.324

Wu, J. (2020) *Demand Forecasting,* [Lectures 3 and 5] Supply Chain Analytics, Imperial College Business School, 26th February and 4th March.

Appendix

Picture 1: Map of Aomori city in Japan.

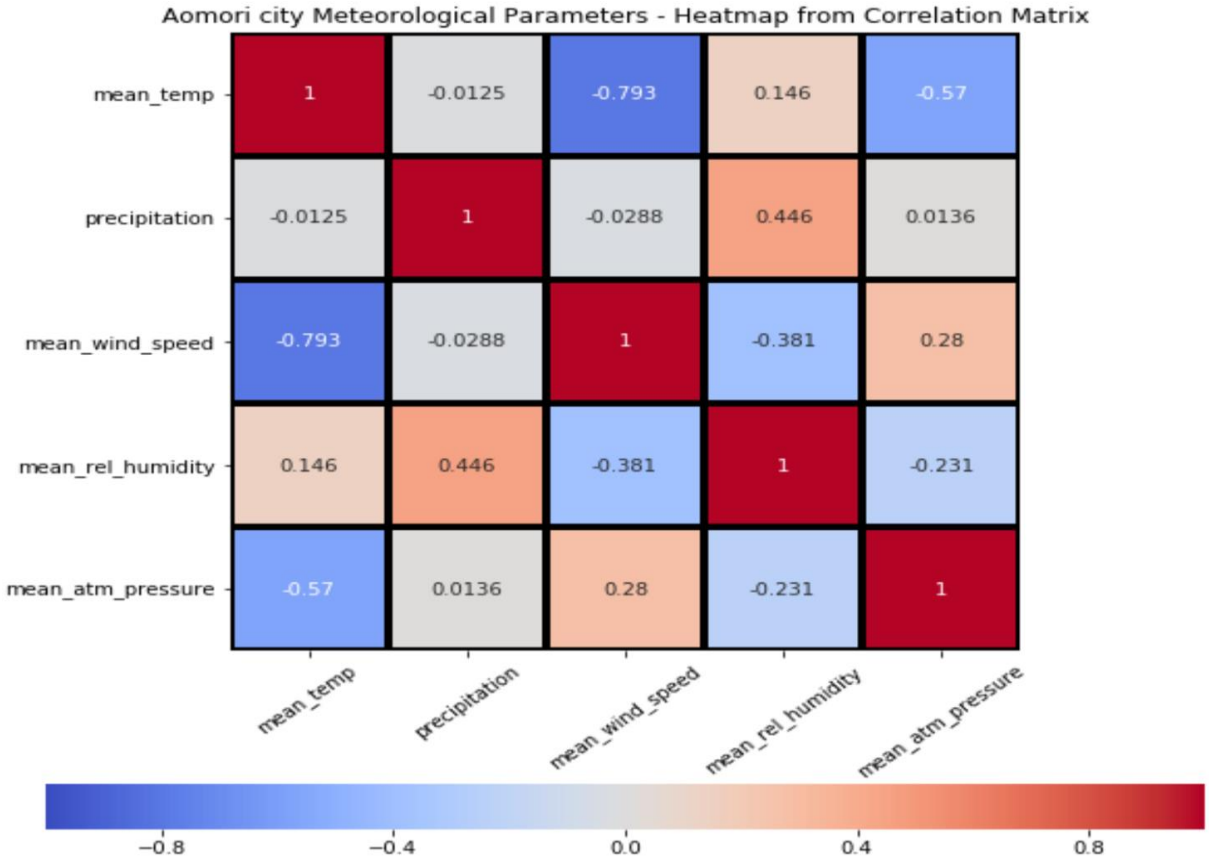

Figure 1: Heatmap visualisation of Correlation Matrix regarding mean temperature, precipitation, mean wind speed, mean relative humidity and mean atmospheric pressure in Aomori city of Japan from 2001 to 2018 by month.

	mean_temp	mean_max_temp	mean_min_temp	max_temp	min_temp	precipitation	highest_daily_prec	mean_wind_speed	highest_wind_gust	mean_rel_humidity	mean_atm_pressure
count	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000
mean	10.675926	14.825463	7.112500	22.119907	2.105093	117.946759	31.150463	13.342130	80.656019	74.740741	1012.865278
std	8.615188	9.187871	8.373934	9.091791	8.174277	62.382546	25.928377	2.095562	15.908107	5.180231	3.468053
min	-2.700000	-0.200000	-5.300000	3.400000	-10.900000	9.000000	4.500000	9.000000	44.300000	55.000000	1005.800000
25%	2.075000	5.825000	-1.200000	14.075000	-5.525000	72.375000	15.375000	11.800000	69.725000	72.000000	1009.900000
50%	11.150000	16.400000	6.850000	23.950000	0.850000	107.000000	23.750000	13.300000	79.600000	75.000000	1013.150000
75%	18.425000	23.125000	14.800000	30.100000	9.225000	148.500000	35.625000	15.100000	89.700000	79.000000	1015.500000
max	26.000000	30.600000	22.200000	36.600000	18.000000	345.000000	208.000000	18.700000	130.700000	85.000000	1021.300000

Table 1: Main descriptive statistics of the data set.

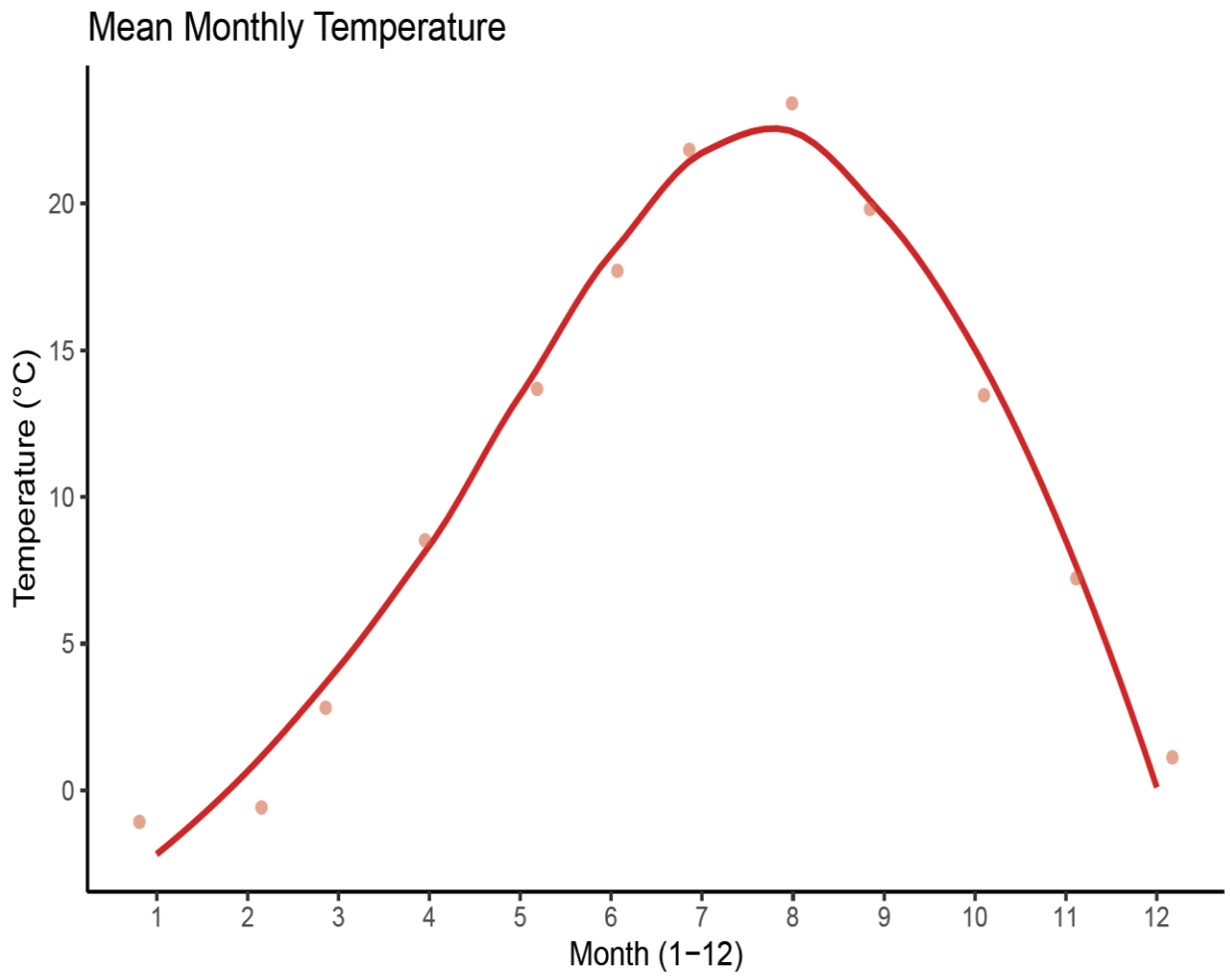

Figure 2: Mean temperature in Aomori city of Japan from 2001 to 2018 by month.

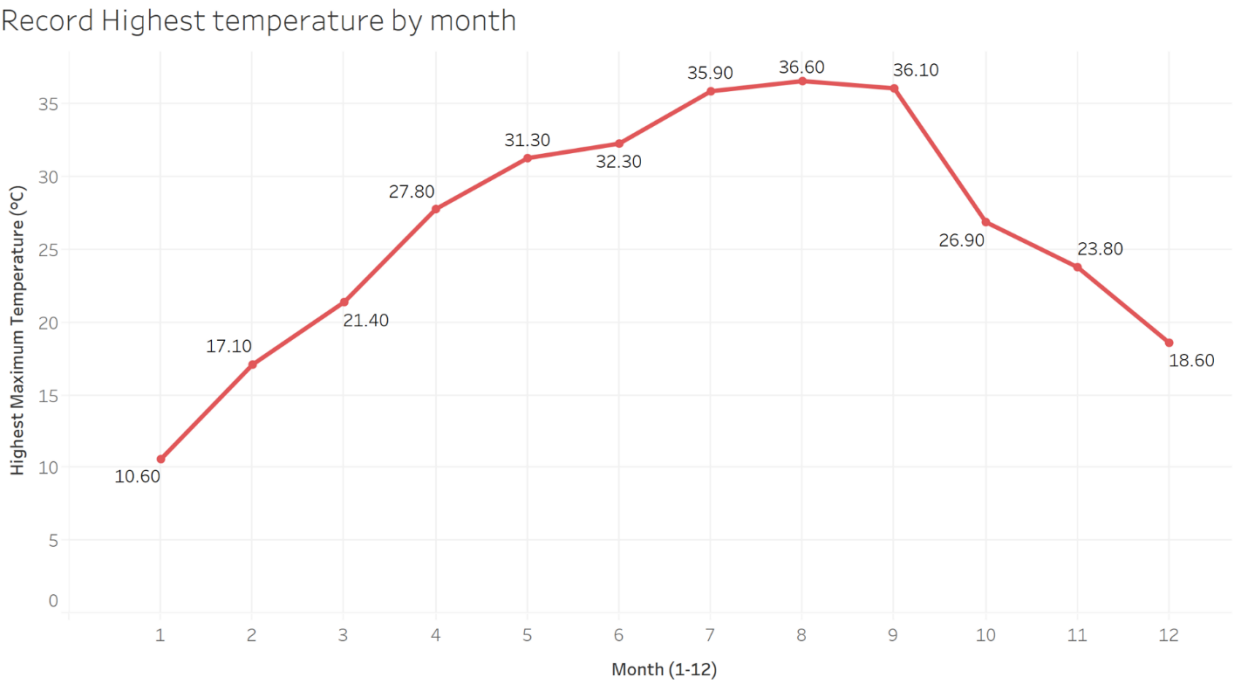

Figure 4: Highest temperature recorded in Aomori city of Japan from 2001 to 2018 by month.

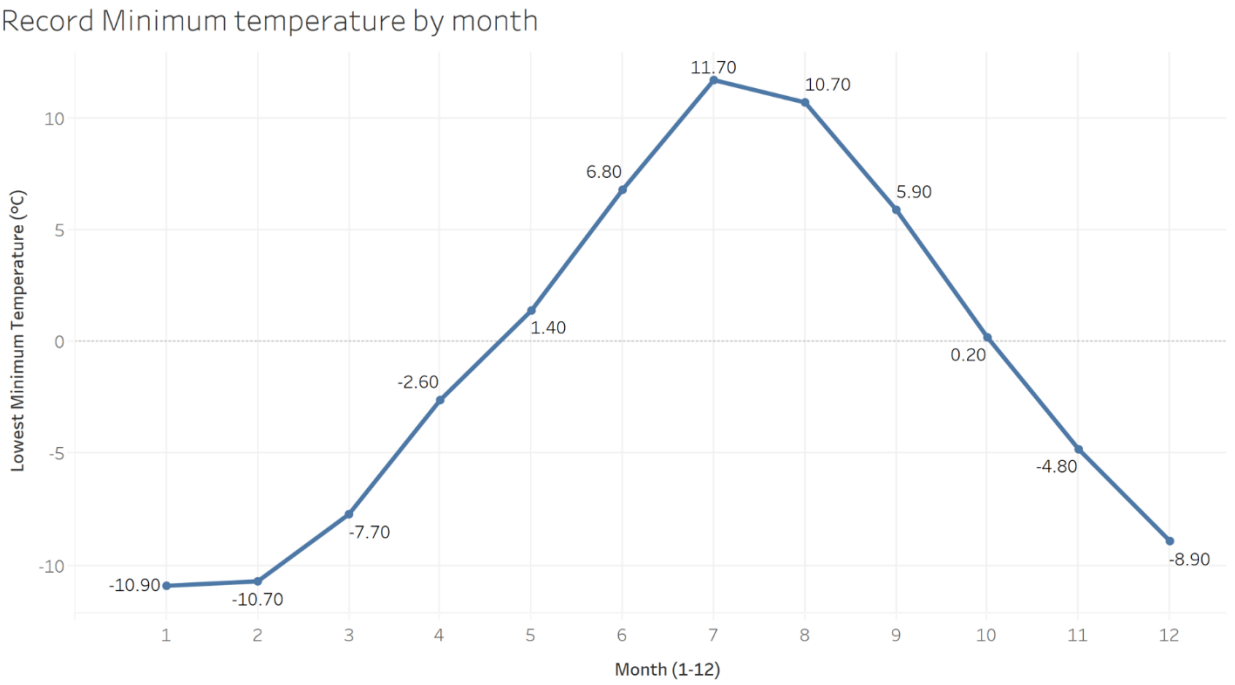

Figure 5: Lowest temperature recorded in Aomori city of Japan from 2001 to 2018 by month.

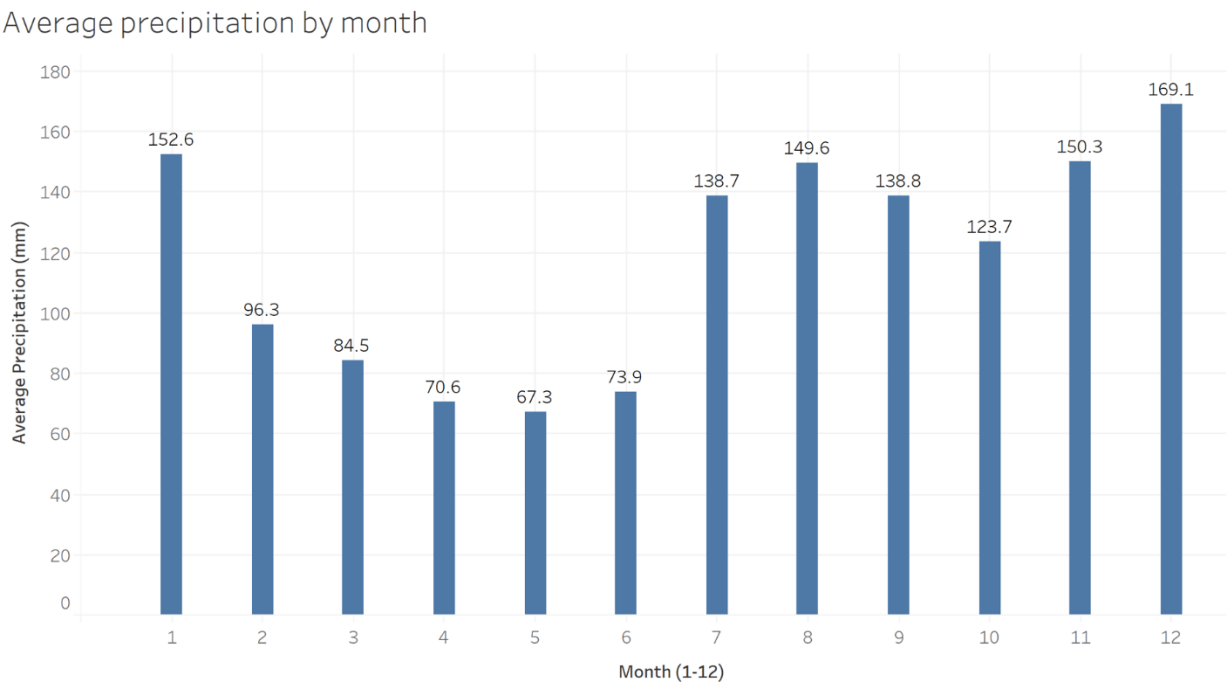

Figure 6: Mean precipitation in Aomori city of Japan from 2001 to 2018 by month.

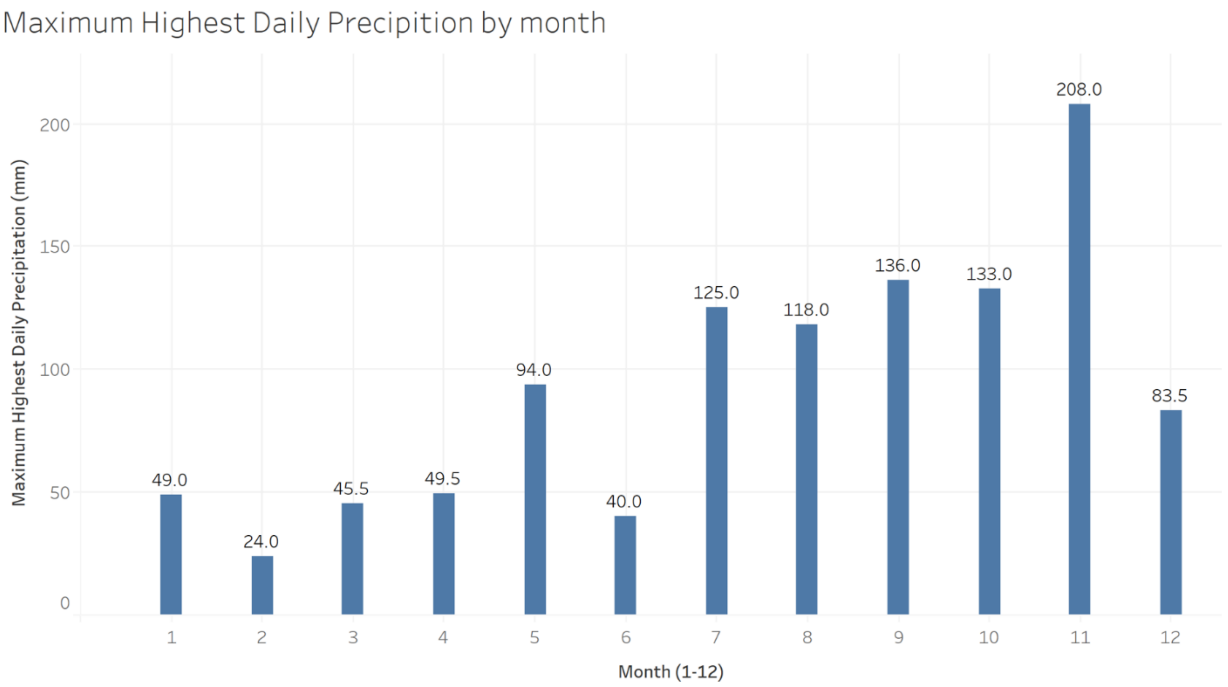

Figure 7: Maximum Highest Daily precipitation in Aomori city of Japan from 2001 to 2018 by month.

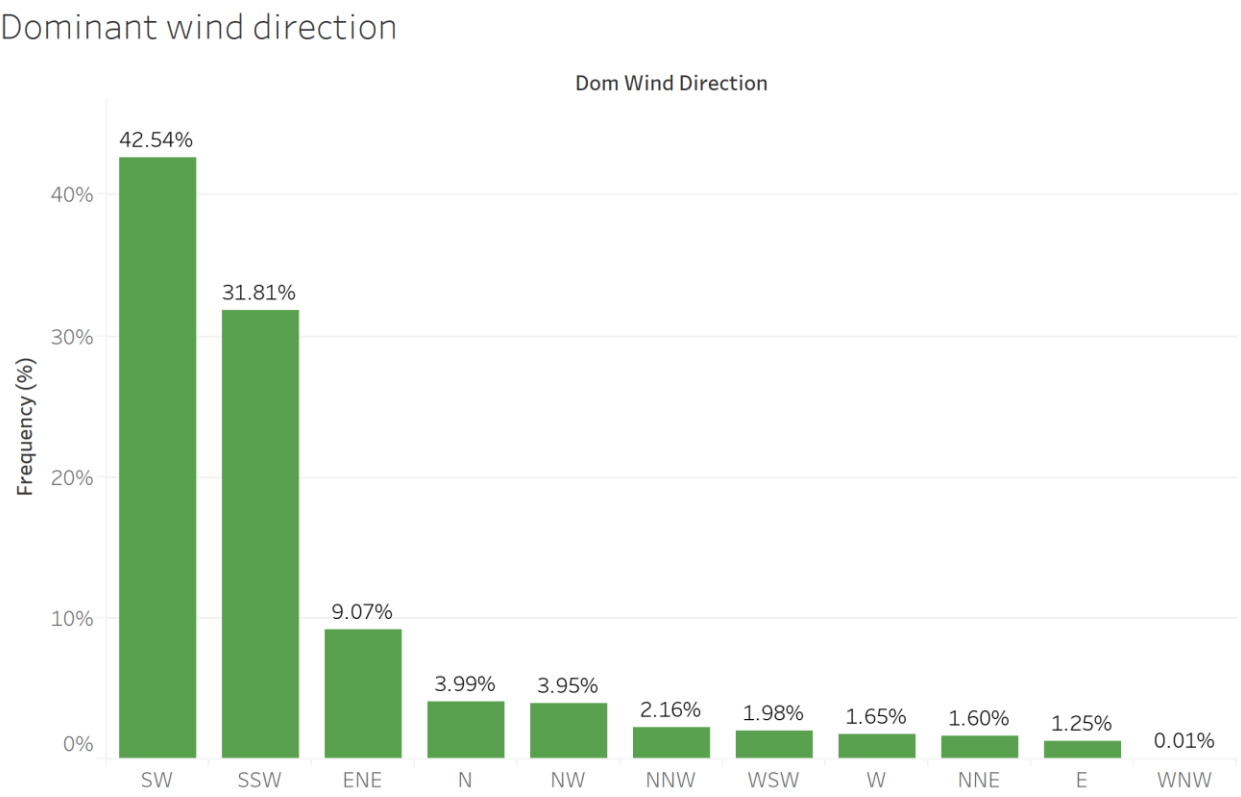

Figure 8: Dominant wind direction in Aomori city of Japan from 2001 to 2018.

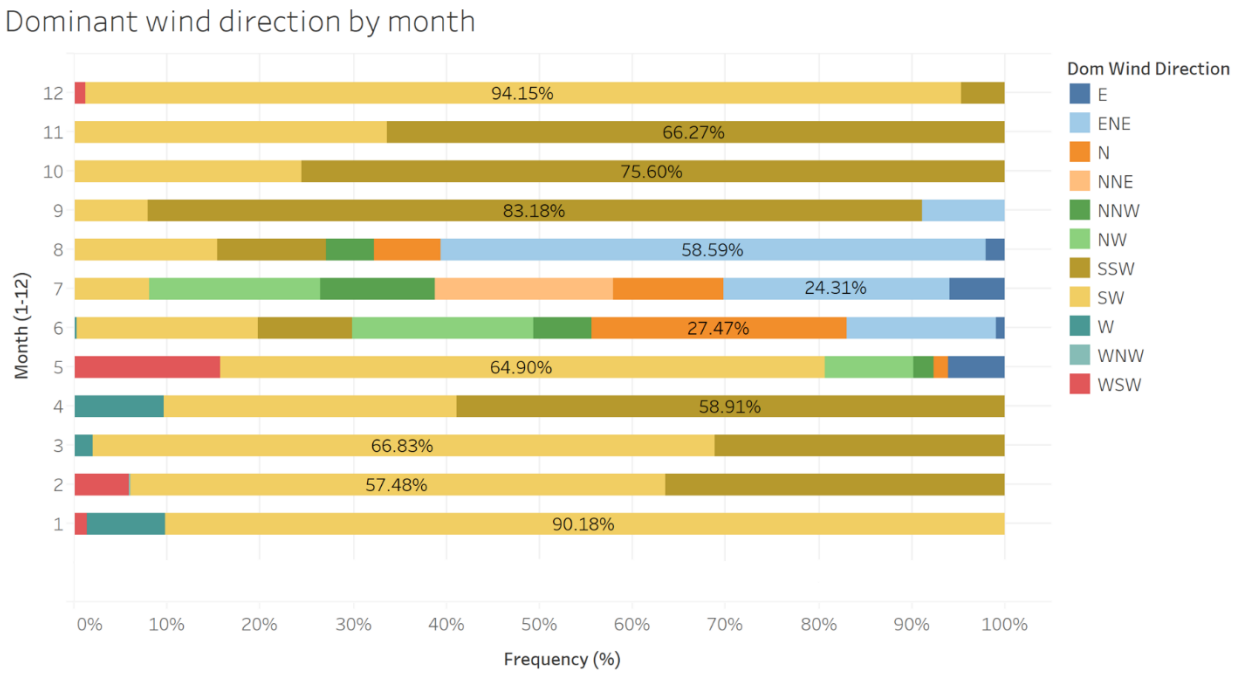

Figure 9: Dominant wind direction in Aomori city of Japan from 2001 to 2018 by month.

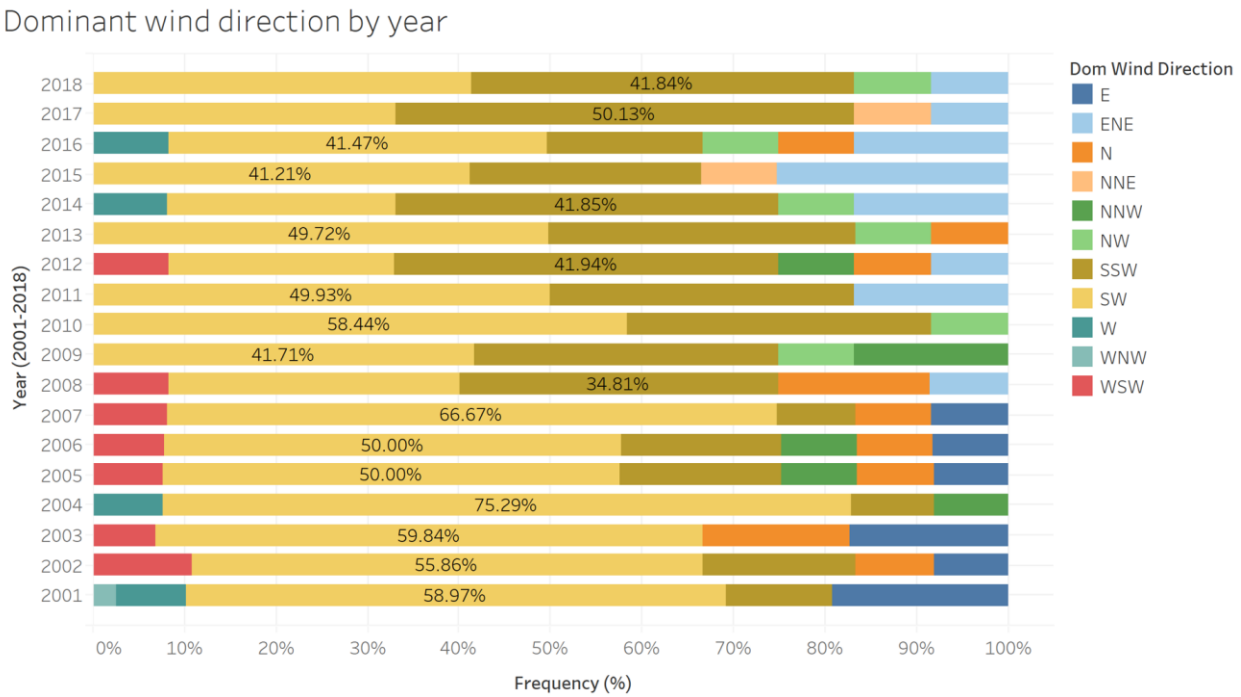

Figure 10: Dominant wind direction in Aomori city of Japan from 2001 to 2018 by year.

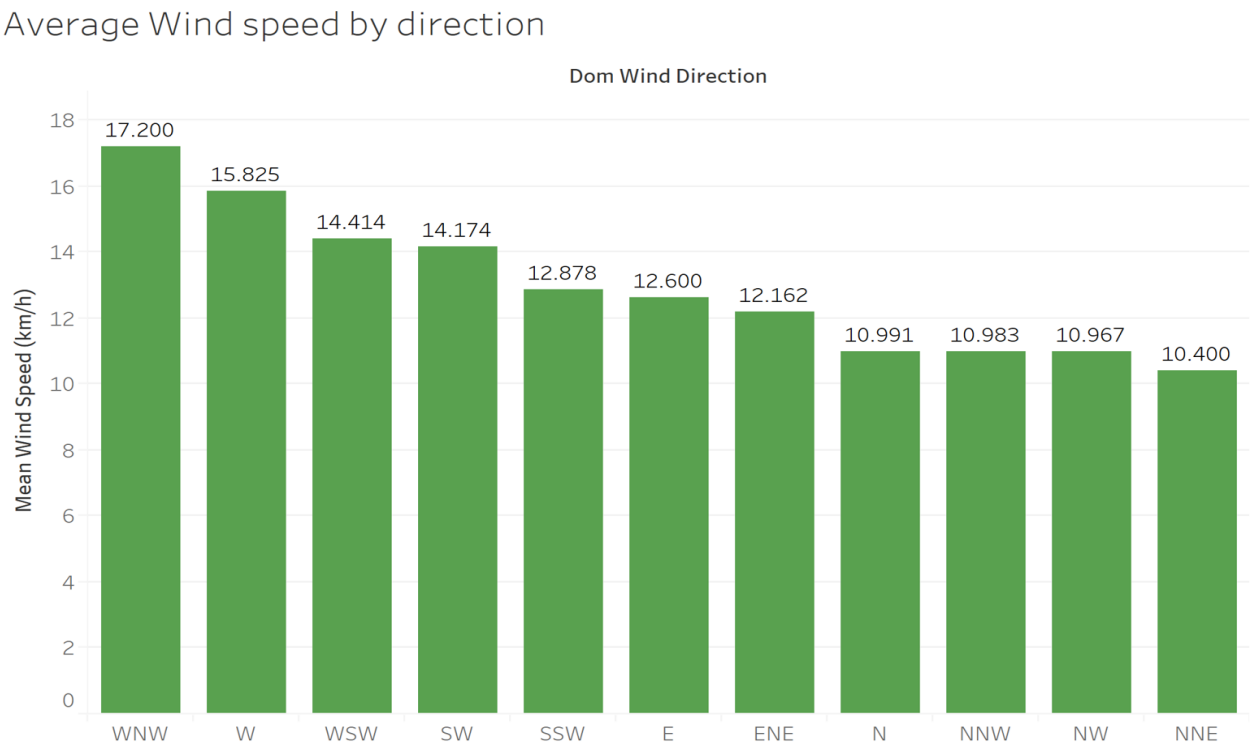

Figure 11: Average wind speed (km/h) in Aomori city of Japan from 2001 to 2018 by wind direction.

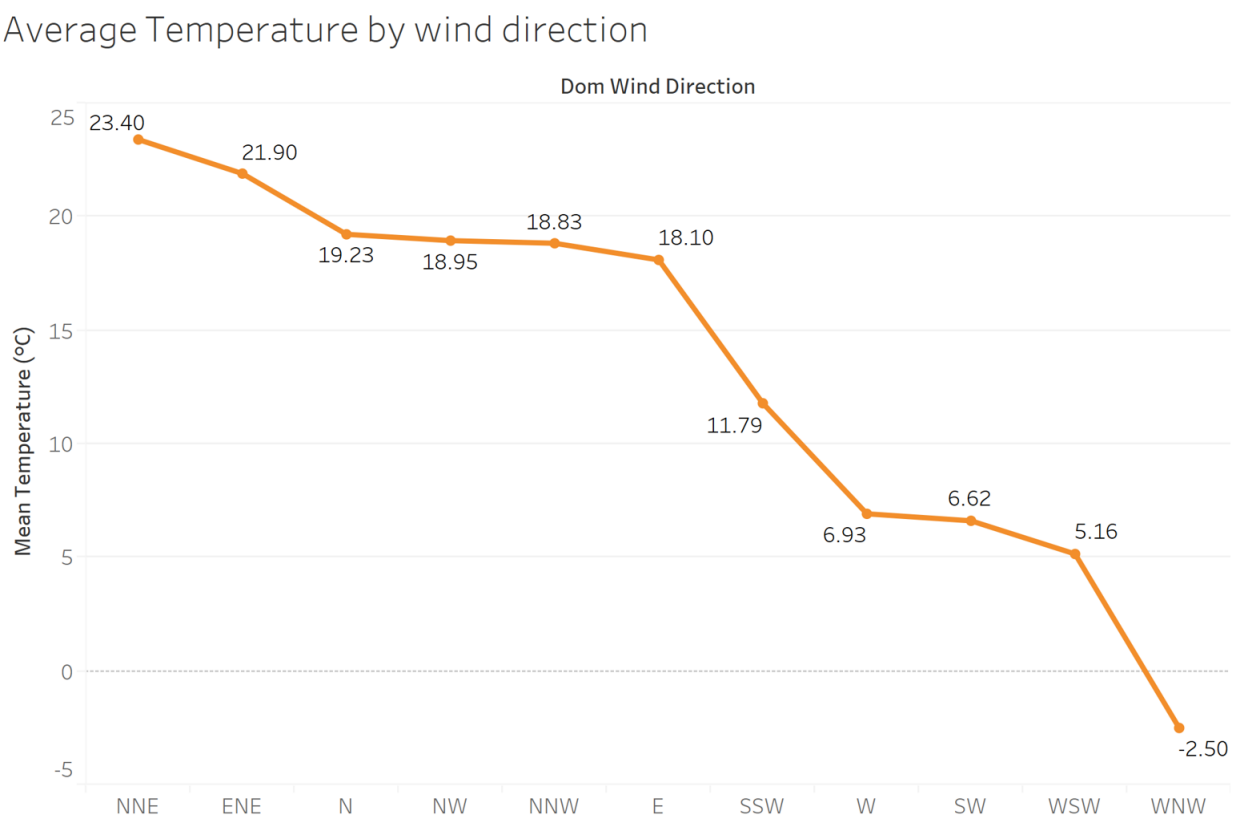

Figure 12: Mean Temperature in Aomori city of Japan from 2001 to 2018 by wind direction.

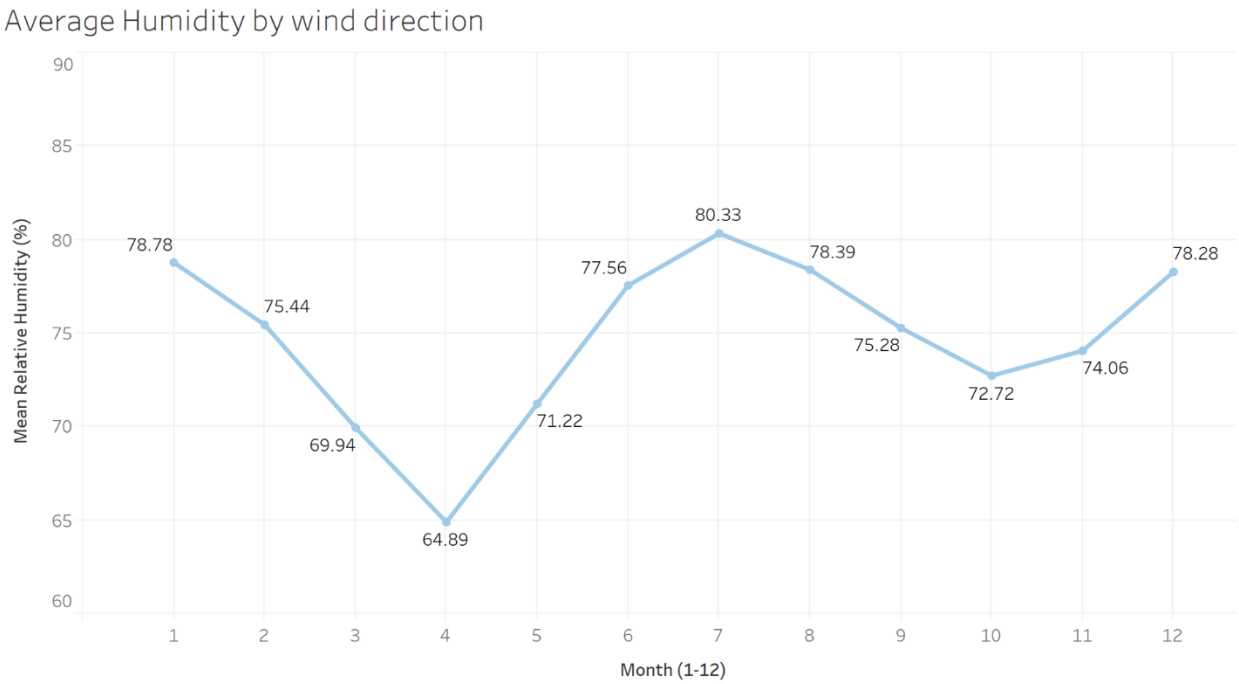

Figure 13: Mean Relative Humidity in Aomori city of Japan from 2001 to 2018 by month.

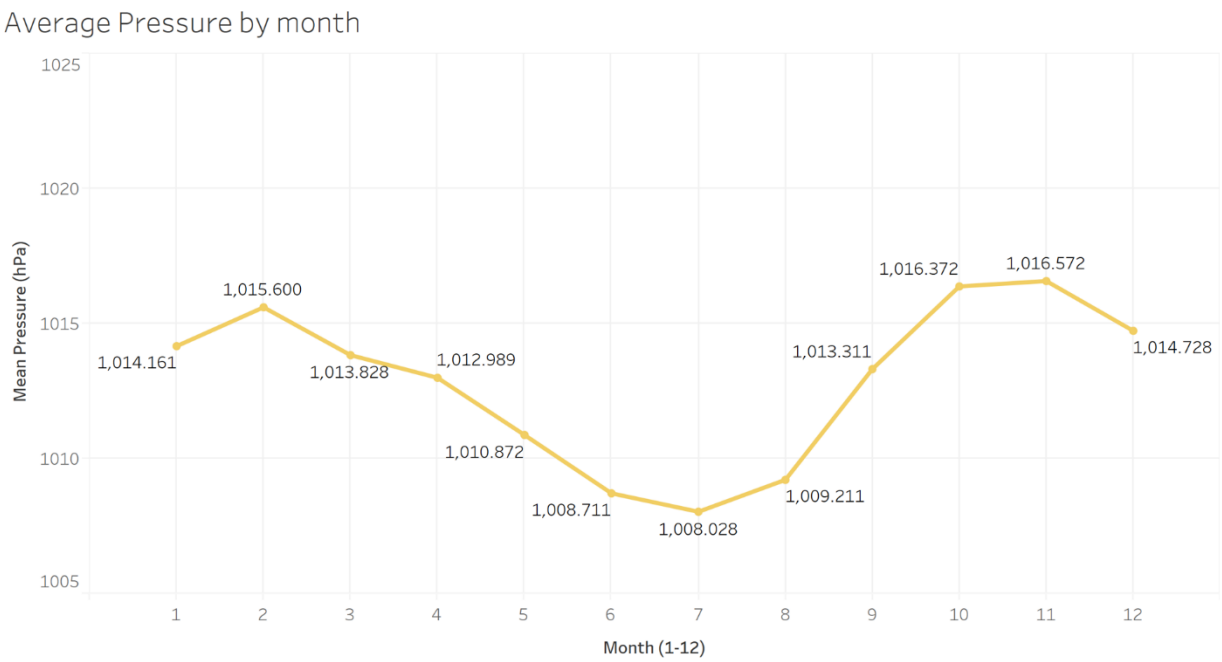

Figure 14: Mean Atmospheric Pressure (hPa) in Aomori city of Japan from 2001 to 2018 by month.

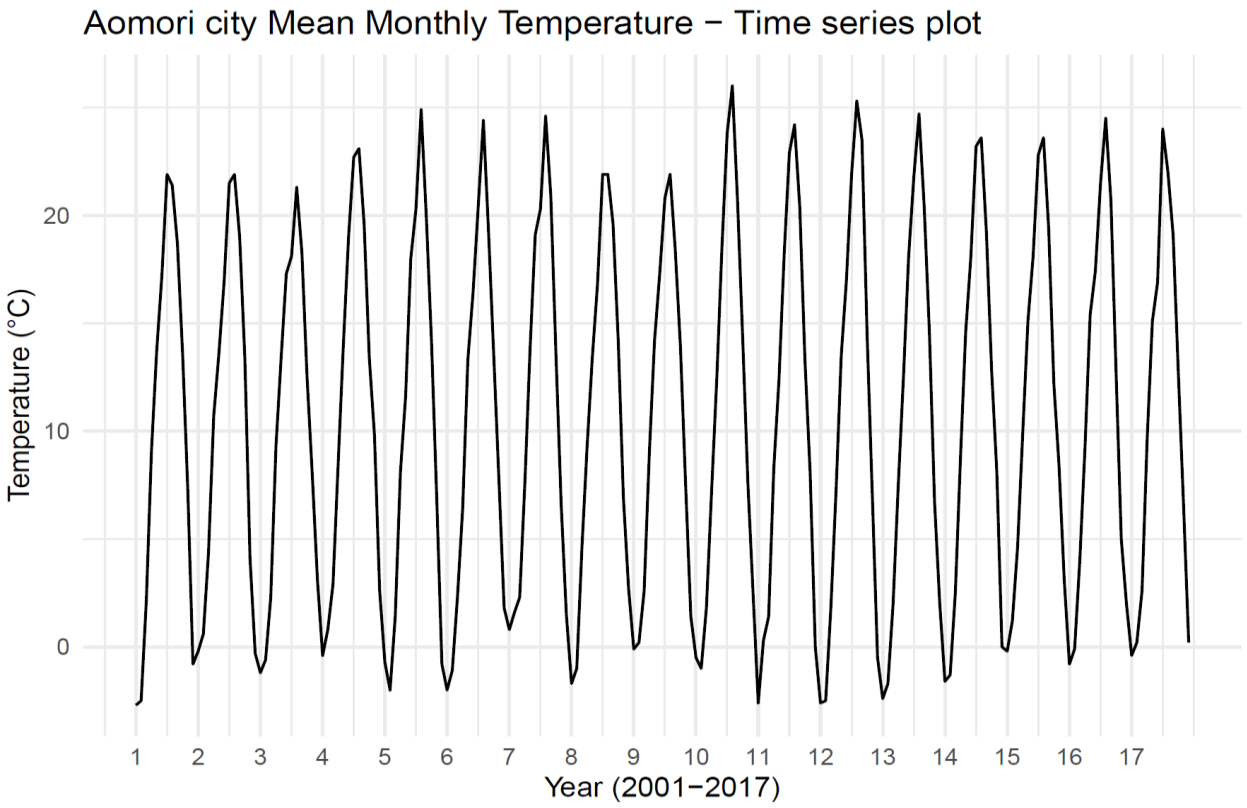

Figure 15: Mean monthly temperature time series data in Aomori city of Japan from 2001 to 2017 by year.

Null Hypothesis: Time series data is stationary	
KPSS Test statistic	0.0497
1% Significance level	0.7390
5% Significance level	0.4630
10% Significance level	0.3470

Table 2: KPSS Test for stationarity level.

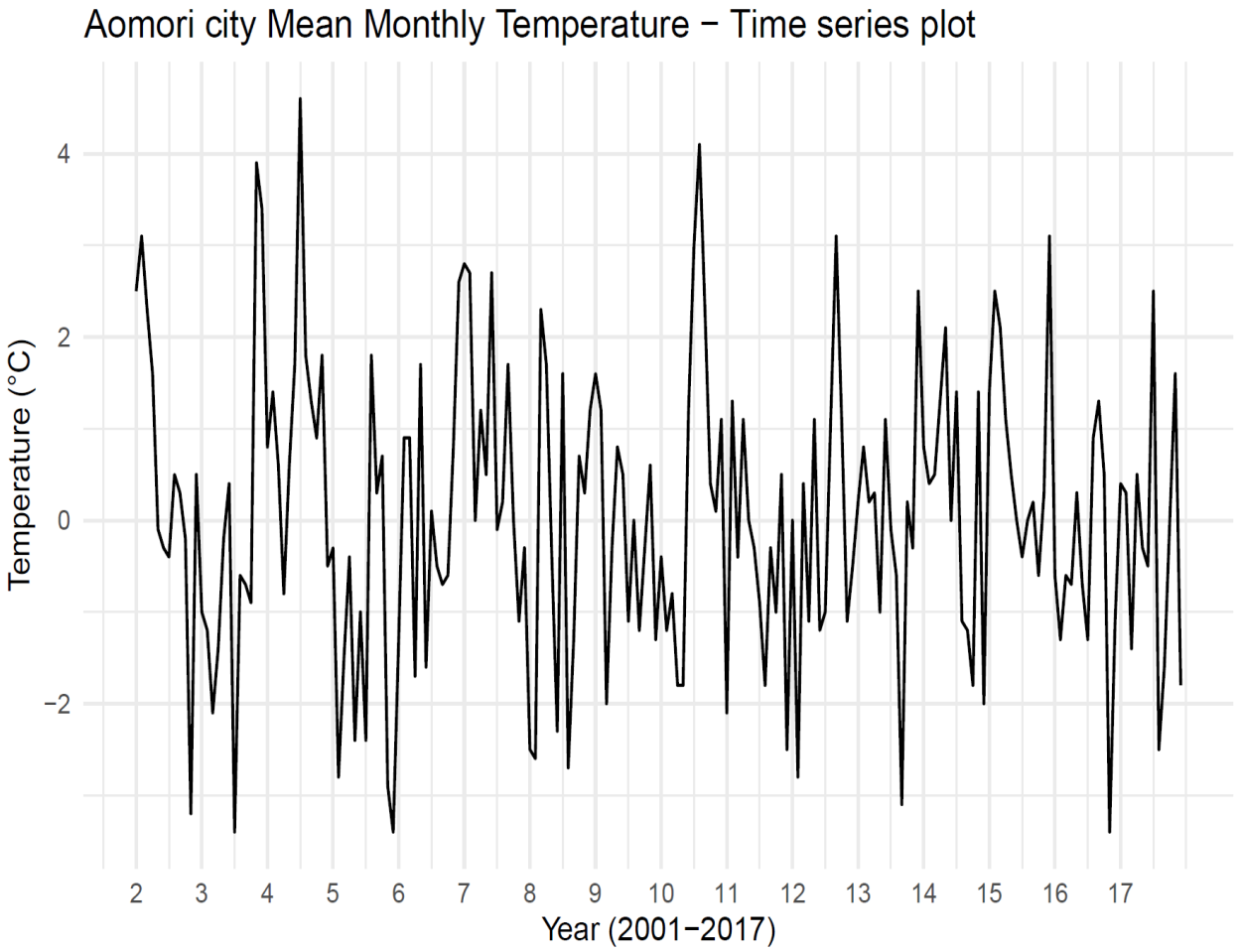

Figure 17: Mean monthly temperature differentiated time series data in Aomori city of Japan from 2001 to 2017 by year.

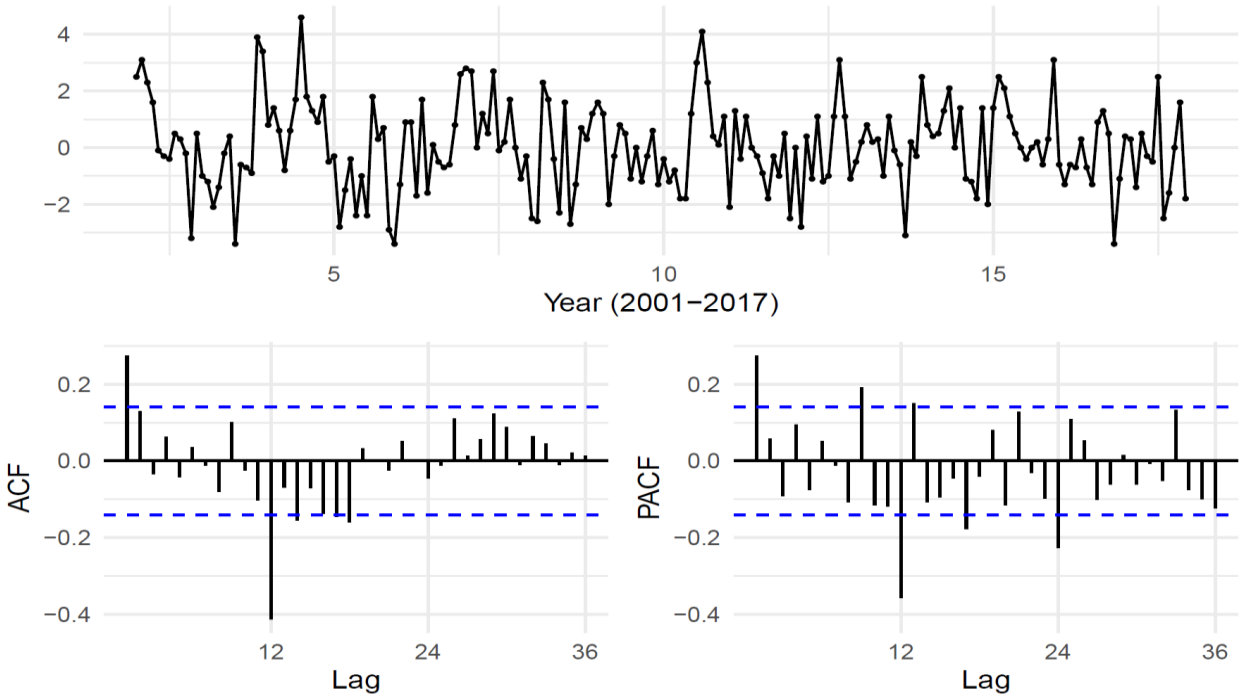

Figure 18: Mean monthly temperature differentiated time series data from 2001 to 2017 by year and Autocorrelation and Partial Autocorrelation function for white noise series in Aomori city.

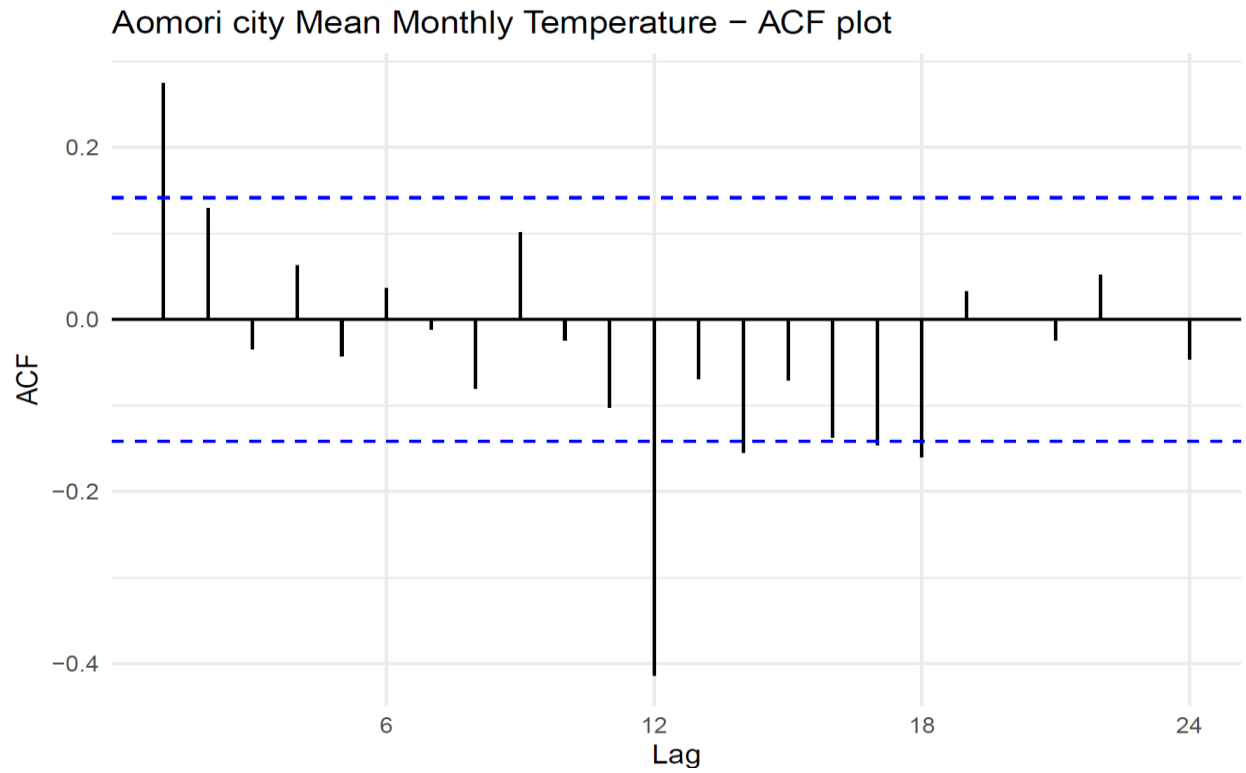

Figure 19: Autocorrelation function for white noise series time-series data in Aomori city.

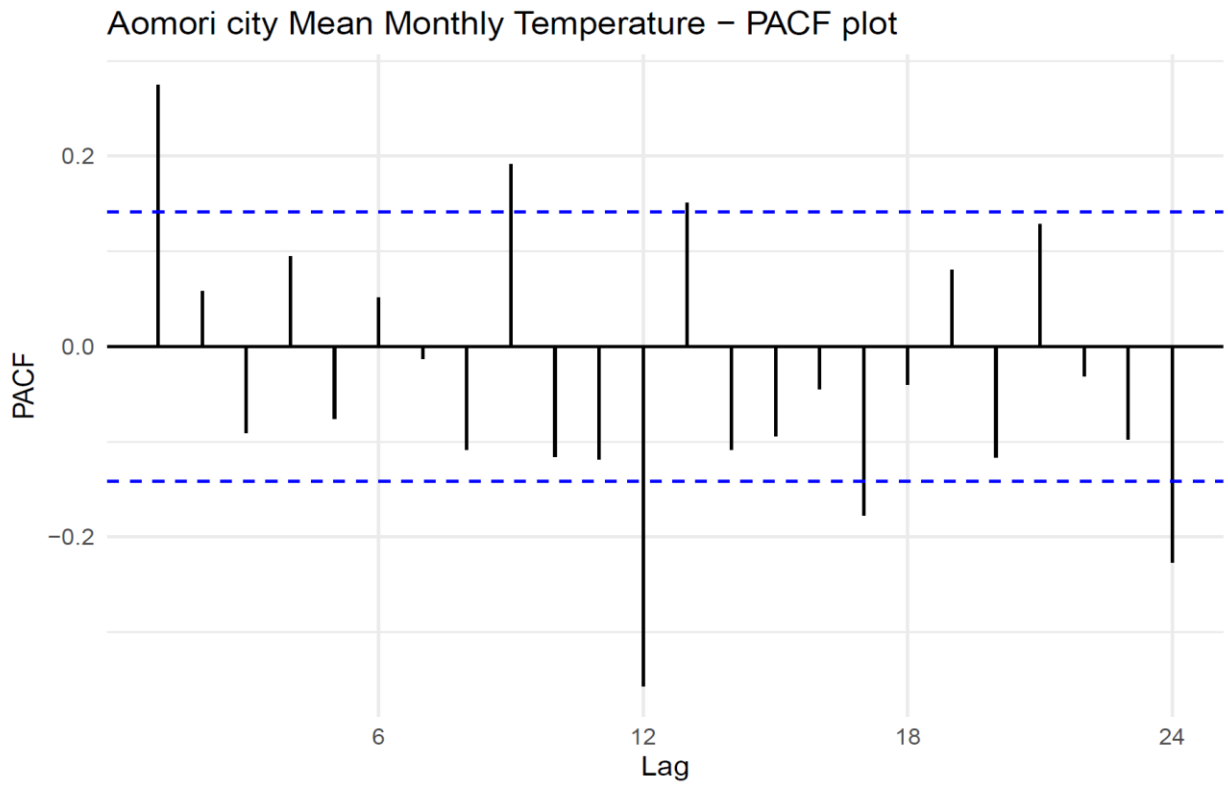

Figure 20: Partial Autocorrelation function for white noise series time-series data in Aomori city.

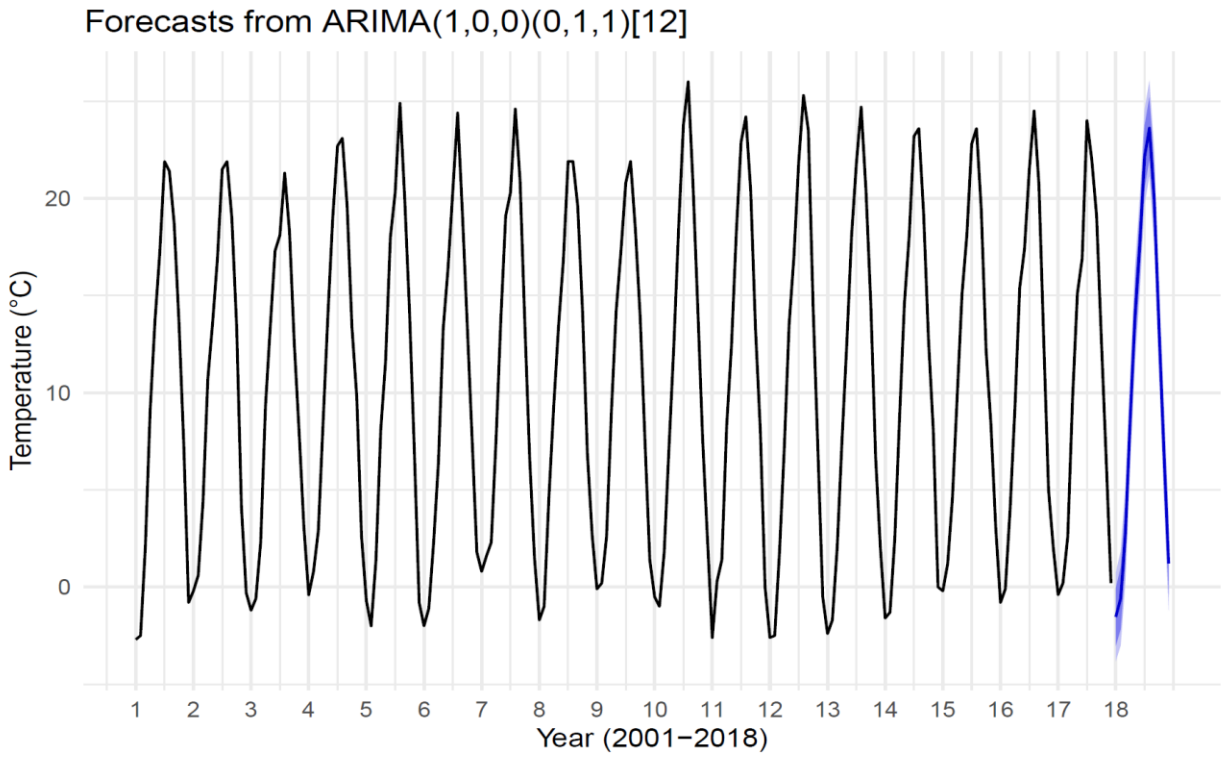

Figure 22: In-sample forecast from Arima(1,0,0)(0,1,1) model.

Month (2019)	Arima Forecast	Lo 80	Hi 80	Lo 95	Hi 95
January	-1.1	-2.6	0.4	-3.3	1.2
February	-0.6	-2.2	0.9	-3.0	1.7
March	3.0	1.4	4.6	0.6	5.4
April	8.8	7.2	10.3	6.4	11.2
May	14.0	12.4	15.5	11.6	16.4
June	17.7	16.0	19.2	15.3	20.1
July	22.2	20.6	23.8	19.8	24.6
August	23.5	21.9	25.1	21.1	25.9
September	20.0	18.4	21.5	17.5	22.3
October	13.4	11.9	15.0	11.0	15.8
November	7.2	5.7	8.8	4.8	9.6
December	1.2	-0.3	2.8	-1.2	3.6

Table 4: Final Forecast of Arima(1,0,0)(0,1,1) model.

Month (2019)	ETS Forecast	Lo 80	Hi 80	Lo 95	Hi 95
January	-0.7	-2.2	0.7	-3.0	1.5
February	-0.2	-1.7	1.3	-2.4	2.1
March	3.2	1.7	4.7	1.0	5.5
April	9.0	7.5	10.5	6.7	11.3
May	14.1	12.6	15.6	11.8	16.3
June	18.1	16.6	19.6	15.8	20.4
July	22.2	20.7	23.7	19.9	24.5
August	23.9	22.4	25.4	21.6	26.1
September	20.2	18.7	21.7	18.0	22.5
October	13.9	12.4	15.4	11.6	16.2
November	7.7	6.2	9.2	5.4	10.0
December	1.7	0.2	3.2	-0.6	3.9

Table 6: Final Forecast of ETS(A,A,A) model.