ST422 Summative Project

Candidate Number:

Question 1

General Data Description

The ‘Sunshine’ data runs from January 1919 to August 2020; the ‘Temperature’ data runs from January 1884 to August 2020; and the ‘Rainfall’ data runs from January 1862 to August 2020.

At first glance, there are clear cyclical trends to all the datasets, evidenced in figure 1 and 2. However, ‘Rainfall’ appears to be less periodic and more variable than either ‘Sunshine’ or ‘Temperature’ Moreover, there does not appear to be a trend in the datasets as the values in the data set remain variable round a constant mean throughout the sample period.

Figure 1

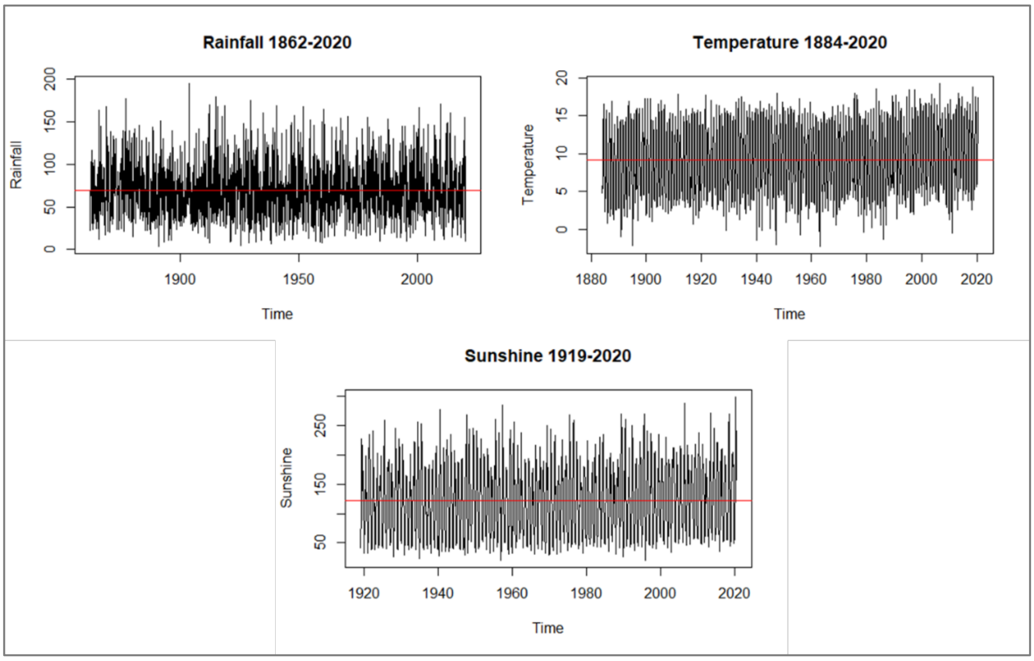
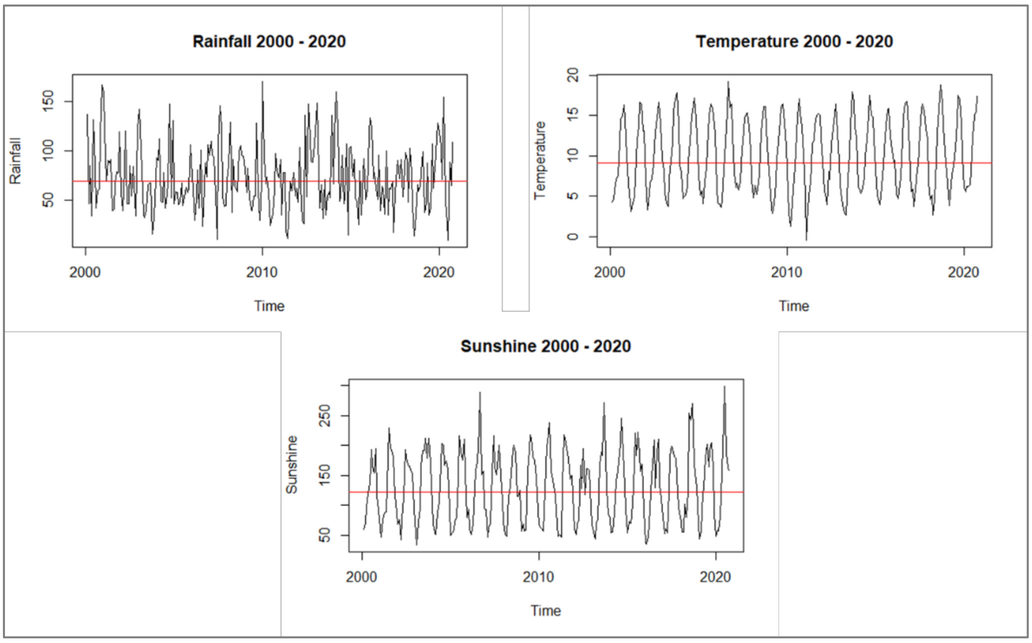


Figure 2



After conducting an ACF analysis on the datasets, it is evident that there is indeed no noticeable trend and the data is cyclical (figure 3) as is evidenced by the lagged variables having a pattern of 1 period length and generally having a correlation greater than the 5% confidence intervals. This indicates that variables which are 1 period apart (12 months) are significantly positively correlated and variables which are ½ period apart (6 months) are significantly negatively correlated. However, as was noticed above, 'Rainfall' is less dependent on seasonal trends as is illustrated with lagged variables only having a slightly greater autocorrelation than the 5% confidence interval. After adjusting for the 12-period seasonality by taking the difference in the data (with a period of 12) the new ACF does not exhibit seasonality (to the same extent) as the unadjusted data (figure 4). However, there does still appear to be a relationship at lag 1. This is due to artificial seasonality generated by finding the difference of period 12 in the data.

Figure 3

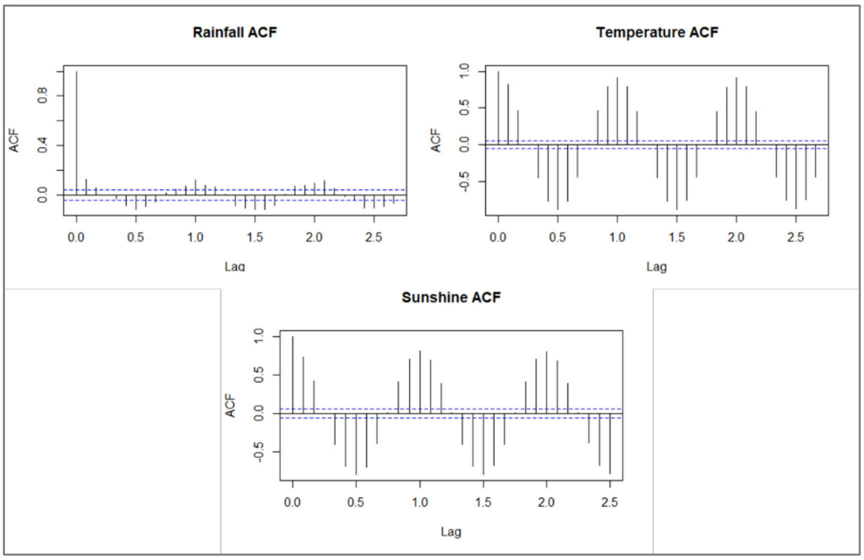
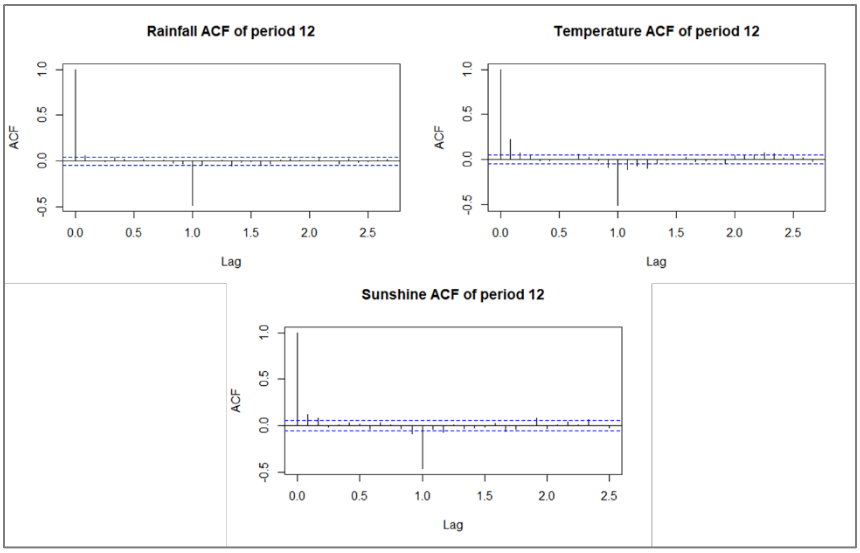


Figure 4



Question 2

Due to the evident seasonality within each of the datasets and the lack of a trend, one would assume that using a SARIMA model similar to the below would capture the seasonality and time dependence of the data in each dataset:

$$x_t \sim ARIMA(p, d, q) \times (P, D, Q)_{12}$$

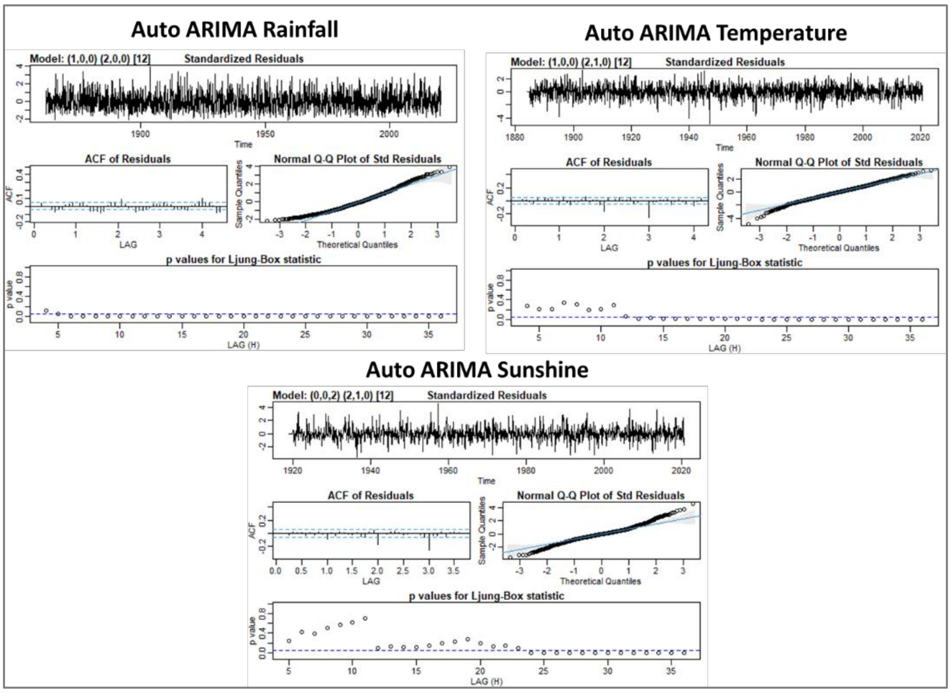
With: $d = 0$ & $D > 0$

Initially the **auto.arima()** function from the 'forecast' package in R was used to fit the initial SAMRIA models for each dataset. The fitted outputs were as follows:

	Rainfall	Temperature	Sunshine
AIC[1]	18,563.09	5,811.23	11,614.12
Fitted Model	$ARIMA(1,0,0) \times (2,0,0)_{12}$	$ARIMA(1,0,0) \times (2,1,0)_{12}$	$ARIMA(0,0,2) \times (2,1,0)_{12}$

Models fitted with the **auto.arima()** function are not always optimal and can be improved. This is evidenced by the fitted model for 'Rainfall' not accounting for seasonality ($D = 0$) and as such the ACF of the residuals exhibit seasonality (Figure 5). Moreover, for the 3 models, the Ljung-Box test showed that lags (particularly later lags) were below a p-value of 5% (indicating that residuals had autocorrelation and thus were not likely i.i.d) for all models fitted and the absolute value of some residuals were greater than 4 indicating that the residuals are not likely to be normally distributed.

Figure 5



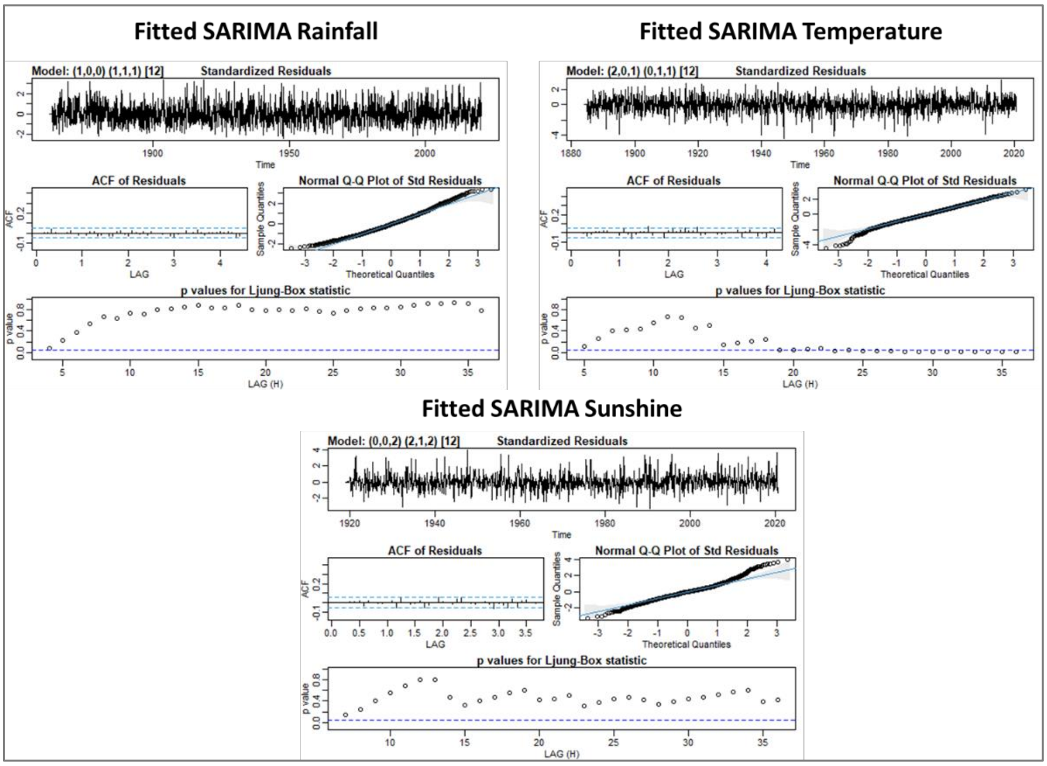
[1] Throughout the report, AIC is used to determine the most suitable model when making comparisons across models, unless otherwise stated. This is primarily because the AIC takes into account both complexity and accuracy of a certain model. Models with more parameters will always have a better fit to their training data, however, they are susceptible to overfitting and being inadequate for real world data. As such the number of parameters in a given model are used as a proxy to adjust for potential overfitting of more complex models. Furthermore, AIC is more widely known than BIC or AICC and thus is more accessible and interpretable to a wider audience.

Therefore, parameters were altered and new models were fitted with **sarima()** function:

	Rainfall	Temperature	Sunshine
AIC	18,314.52	5,396.74	11,298.49
AIC decrease	1.34%	7.13%	2.72%
Fitted Model	$ARIMA(1,0,0) \times (1,1,1)_{12}$	$ARIMA(2,0,1) \times (0,1,1)_{12}$	$ARIMA(0,0,2) \times (2,1,2)_{12}$

The above models were the 'best' models achieved adjusting the hyperparameters (p, d, q, P, D, Q). They all have lower AIC scores than the auto fitted models and they all consider seasonality (by design). Further, the lags for the Ljung-Box test are significant for up to 35 lags in 'Rainfall' and 'Sunshine' and have improved in the case of 'Temperature'. Moreover, the absolute value of residuals has decreased in the new 'Rainfall' model. However, the residuals of temperature are still greater absolutely than 4 in some cases and not all the Ljung-box lags are significant. This suggests that the SARIMA models fitted for temperature may not be sufficient to model temperature (figure 6).

<u>Figure 6</u>



Therefore, due to the more favourable AIC scores, improved residuals and improved Ljung-Box Static test, the adjusted ARIMA models were used to model the datasets. The models were as follows:

Rainfall: $x_t \sim ARIMA(1,0,0) \times (1,1,1)_{12}$

Temperature: $x_t \sim ARIMA(2,0,1) \times (0,1,1)_{12}$

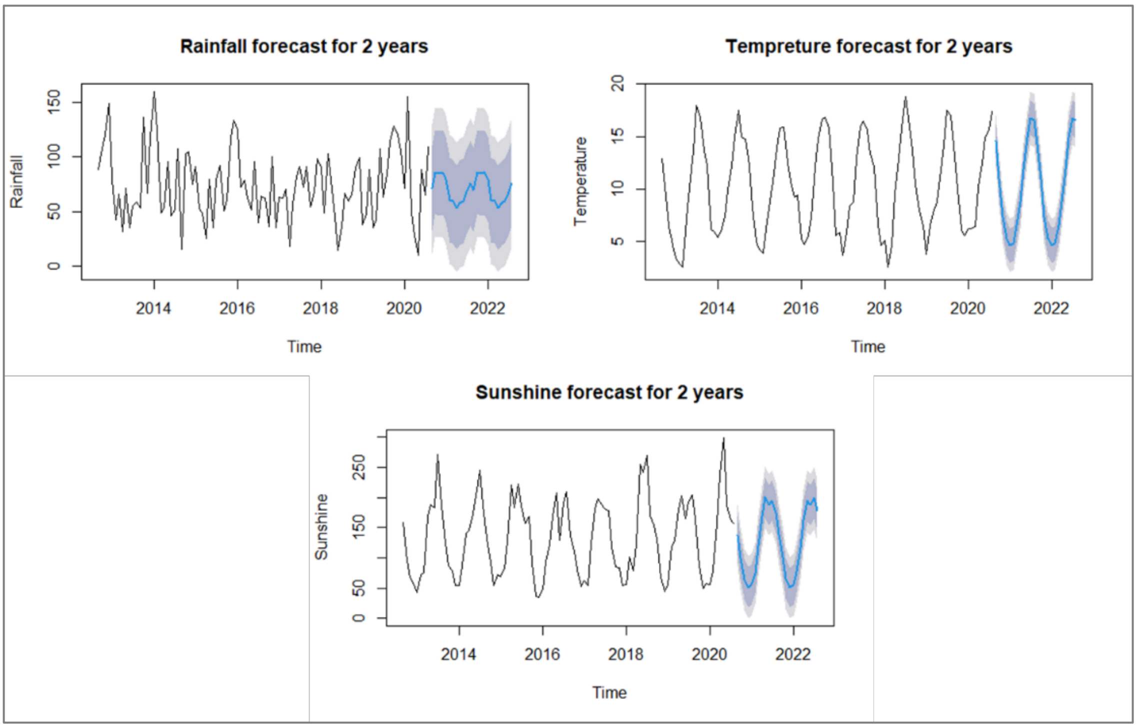
Sunshine: $x_t \sim ARIMA(0,0,2) \times (2,1,2)_{12}$

Using the above models and the whole datasets, the mean forecasted measurements for November and December were:

	November	December
Rainfall	85.01039	85.52339
Temperature	7.330432	5.297409
Sunshine	64.70617	51.77630

Further, figure 7 shows the forecasted mean and confidence intervals for the coming 2 years for each data set using the chosen models. It is interesting to note that although the validity of the 'Temperature' model (based on absolute value of residuals and Ljung-Box static) is questionable, the forecast confidence intervals of the model are very small for all months except 'extreme' months i.e mid-summer or mid-winter. This is also a characteristic of the 'Sunshine' forecasted values. This characteristic suggests that the maximum and minimum temperature/sunshine is more variable than the intervening months mean values. Consequently, the extreme months are responsible for the large residuals in the fitted models and drive the error rate in the models. Unlike 'Sunshine' and 'Temperature', 'Rainfall' is variable in all periods of the year and the forecasted confidence intervals indicate that rainfall is difficult to predict. Thus, once the slight seasonality of rainfall is adjusted for, the errors are less likely to still exhibit autocorrelation (Ljung-box static) and are more likely to appear as i.i.d normally distributed errors.

Figure 7



Question 3

The datasets were split into training (first 75% of observations) and validation data (last 25% of observations). Multiple rolling window approaches with different window lengths (M) and recalculations periods (A) were implemented to forecast the next monthly measurement in the training and validation data. At each recalculation period, an **auto.arima()** was fitted to the past M observations. This fitted model was used for the following 'A' observations to forecast a subsequent months measurement based on the past 'M' observations.

The forecasted monthly measurements were compared to the actual measurement and the monthly mean squared error was computed which when aggregated across all the monthly predictions gave the 'error' for each of the rolling window models implemented.

Below are 3 tables summarizing the results of rolling windows with different window lengths and recalculation periods (M & A) on the training data for 'Rainfall', 'Temperature' and 'Sunshine':

Squared Monthly Prediction Error – **Rainfall** (training data)							
M	60	120	180	180	180	**210**	500
A	36	36	36	72	108	**108**	108
Error	1072.997	1062.631	1036.658	1040.594	1034.904	**1020.735**	1039.349
$Mean^2$	4,799.56						
Normalised Error	0.223562	0.221402	0.21599	0.21681	0.215625	**0.212673**	0.216551

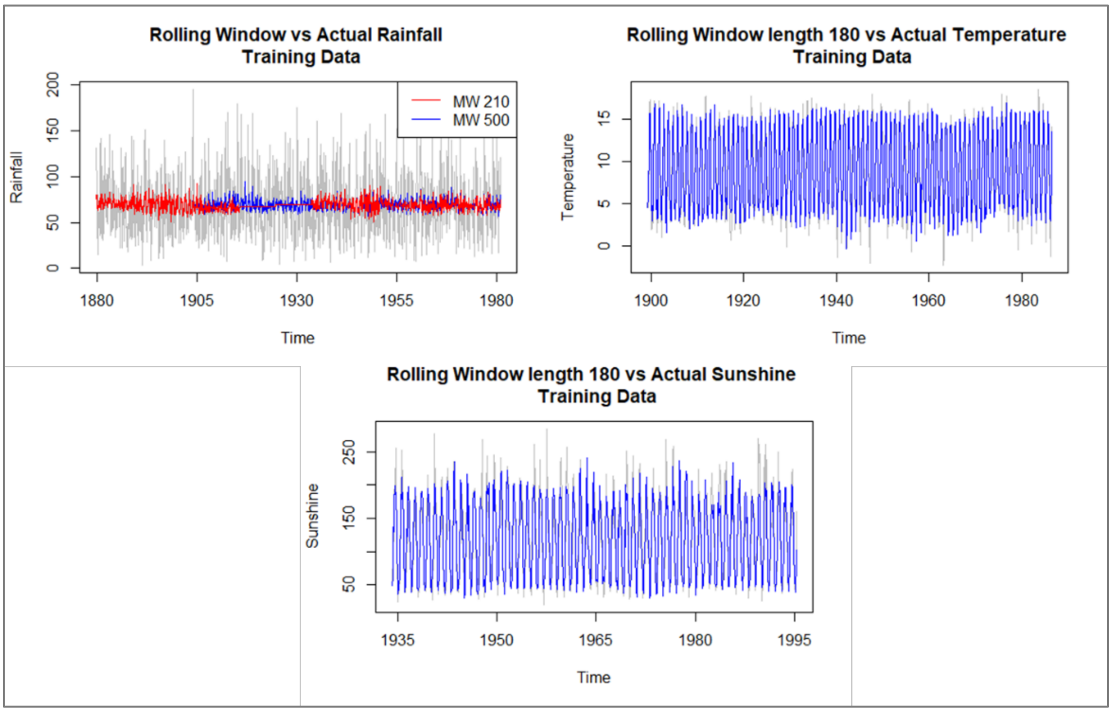
Squared Monthly Prediction Error - **Temperature** (training data)							
M	60	30	60	90	120	150	**180**
A	36	72	72	72	72	72	**72**
Error	2.544401	3.785896	2.465215	2.313719	2.104553	2.090774	**2.041671**
$Mean^2$	79.031						
Normalised Error	0.032195	0.047904	0.031193	0.029276	0.026629	0.026455	**0.025834**

Squared Monthly Prediction Error – **Sunshine** (training data)						
M	60	60	60	60	90	**180**
A	36	72	108	144	108	**108**
Error	1032.664	997.5507	973.7345	1036.108	865.7867	**838.5785**
$Mean^2$	14,212.3116					
Normalised Error	0.07266	0.070189	0.068513	0.072902	0.060918	**0.059004**

Firstly, the 'normalised error' indicates adjusts the error for the scale of the measurements of an individual dataset. As such, we can see which models have a relatively lower/higher error. 'Temperature' and 'Sunshine' both have low normalised errors indicating that they are quite predictable. However, the normalised error for 'Rainfall' is almost 1 order of magnitude greater than 'Temperature' normalised errors, indicating that 'Rainfall' is difficult to predict.

Theoretically a model with a larger M, given the same A will always have a lower error for the same observations. However, this is not the case in 'Rainfall' because when a model's window length is larger, there are less observations in a given sample to estimate (the first M observations have to be skipped). Therefore, a rolling window model with a smaller window will have more observations from a sample than a model with a larger widow length. This is illustrated in figure 8. As one can see, a rolling window of M = 210 is applicable from 1880 whilst model with M = 500 is only applicable from 1905. Further, one will notice that from 1905, the model with M = 500 captures more of the variability in rainfall. However, the average mean squared error for M = 210 from 1880 – 1905 will (if smaller than the average mean squared error for 1905 – 1980) bring down the average mean squared error over all the observations ('error'). Thus, it is possible that a smaller window rolling window model can have a lower 'error' than a model with a grater window length.

Figure 8



The 'Sunshine' and 'Temperature' rolling windows 'track' the actual data very well in most months expect for those where values tend to be extreme such as July and December. These observations are in line with those made with the charts in figure 7. Additionally, the 'Rainfall' rolling window does not follow the underlying observations well, which is consistent with the observations made in figure 7 and the comparison of the normalised errors.

The same rolling window models were applied to the validation data to investigate whether the same window length and recalculation frequency were optimal. For 'Rainfall' data, the optimal rolling window had the same A and M as in the training data. However, different rolling window models were found to minimise the error for 'Temperature' and 'Sunshine'. Nevertheless, it must be noted that only the size of the window was changed, not the frequency at which the SARIMA models are refitted. Further based on the observations made above, one can conclude that alternative window lengths were selected because the additional data available for smaller windows had reduced error rate in excess of the additional error rates that smaller window models had when compared to larger windows in the same observational sample.

Squared Monthly Prediction Error – **Rainfall** (validation data)							
M	60	120	180	180	180	**210**	300
A	36	36	36	72	108	**108**	108
Error	1097.011	1079.676	1073.778	1067.832	1061.908	**1042.967**	1058.067
$Mean^2$	5,095.77						
Normalised Error	0.215279	0.211877	0.210719	0.209553	0.20839	**0.204673**	0.207636

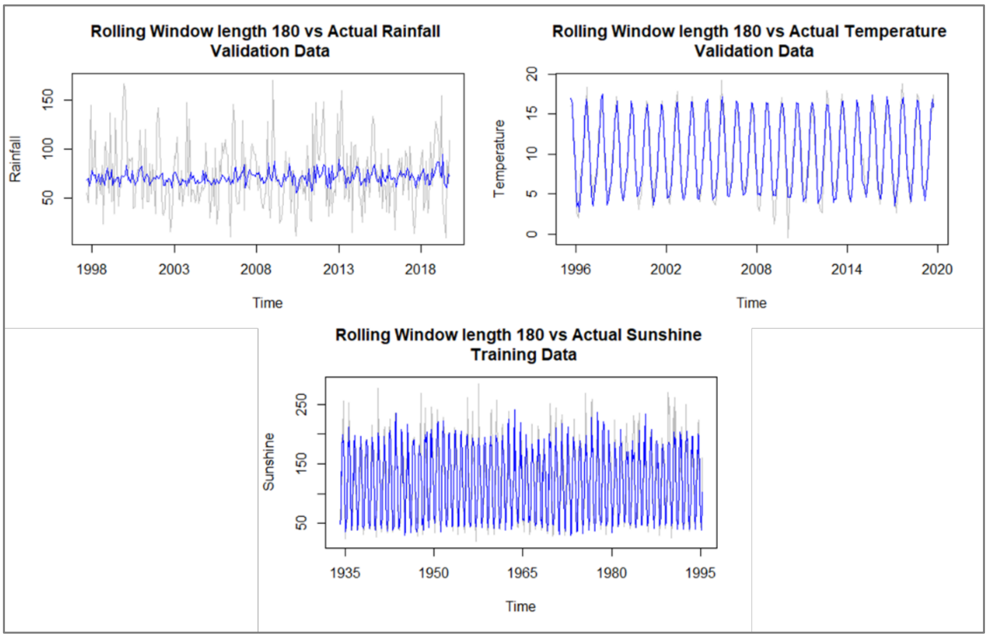
Squared Monthly Prediction Error - **Temperature** (validation data)							
M	60	30	60	90	**120**	150	180
A	36	72	72	72	**72**	72	72
Error	2.124515	3.503016	2.259816	1.801681	**1.588887**	1.720935	1.593192

$Mean^2$	98.19						
Normalised Error	0.021637	0.035676	0.023015	0.018349	**0.016182**	0.017527	0.016226

Squared Monthly Prediction Error – **Sunshine** (validation data)						
M	60	60	**60**	60	90	180
A	36	72	**108**	144	108	108
Error	880.1794	884.6471	**727.1422**	891.6657	828.0783	857.2719
$Mean^2$	17,096.21					
Normalised Error	0.051484	0.051745	**0.042532**	0.052156	0.048436	0.050144

Furthermore, although the optimal rolling window models were different for validation and training data, the error rate of the 'Temperature' optimal training rolling window based on the validation data was not significantly different (1.593192) than that of the error rate from the optimal validation model (1.588887). This suggests that forecasting of 'Temperature' is robust to alternative hyperparameters, which is comforting to observe.

<u>Figure 9</u>

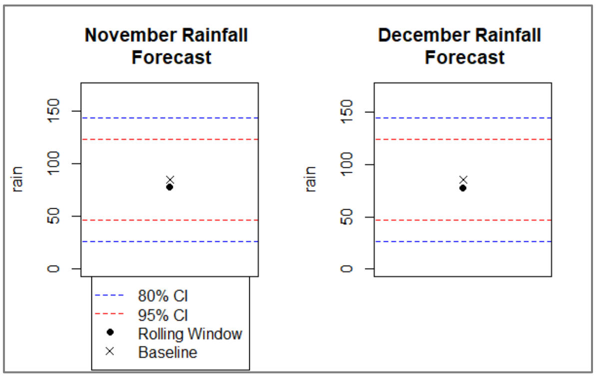


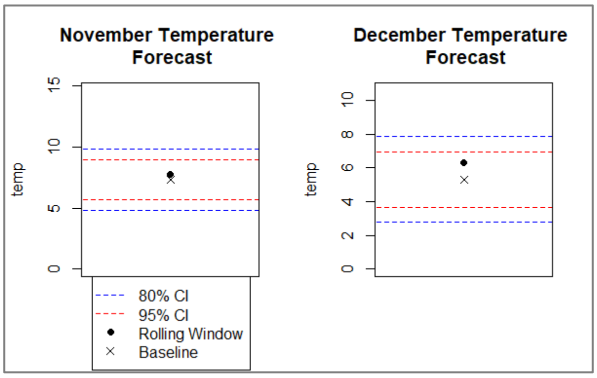
From figure 8 one can see that the optimal 'Rainfall' rolling window for the validation data did not estimate observations well. Furthermore, like the training data, both the 'Temperature' and 'Sunshine' rolling window models predicted the following months observations well in months where extreme measurements were unlikely to occur.

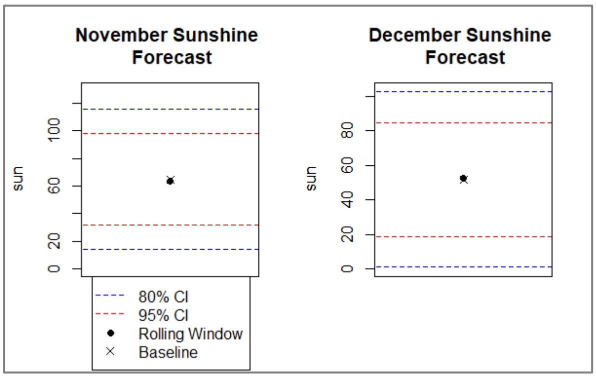
It should be noted that the frequency of refitting does not appear to have a 'simple' interpretable theoretical effect on the 'error' as does the window length. This is evidenced by 'Rainfall' validation data models and 'Temperature' validation models amongst others. In the 'Rainfall' instance, a model with M=180 and A=72 has a lower error than a model with M=180 and A=36, whilst in the 'Temperature' case, a model with M = 60 and A=36 has a lower error than a model with M=60 and A=72. A reasonable explanation is that too frequently refitting the model can lead to a form of overfitting where recent variations increasingly become incorporated into future forecasts.

However, on the other side of the argument, refitting too infrequently can result in the forecasts not incorporating changing conditions within the sample of observations.

I elected to use the SARIMA models that were fitted at the end of the validation data training to forecast the November and December results as they had lower ‘errors’ than the training data models for ‘Temperature’ and ‘Sunshine’. Moreover, the ‘errors’ for the validation ‘Rainfall’ model would have been less than the ‘Rainfall’ training model over the period 1905 – 1980. Further, unlike the ‘Rainfall’ training model, the validation model did not exhibit poor estimation such as the training model between 1915 – 1933. The comparisons of the predicted November and December values for the validation models and the prediction values (mean and confidence intervals) of question 1 are as follows:







The prior figures and remarks have decisively shown that Rainfall is largely unpredictable. As such, forecasts have large confidence intervals. Furthermore, although the forecasts for November and December ‘Rainfall’ are very similar, one cannot be too confident about the predictions given the variability of ‘Rainfall’.

The ‘Sunshine’ and ‘Temperature’ forecasts are reasonable and expected. The rolling window estimates are within 95% confidence interval of the baseline forecast. Moreover, due to the forecast period being at a time of year where extreme values are realised (middle of Winter), the confidence intervals are quite wide (especially in December).

Question 4

A VARMA model can be described as a multivariate ARMA model. Rather than only considering the same time series when forecasting future values of that time series, a VARMA model allows for multiple time series to be used as a basis of forecasting future values of a certain time series. In the case of the datasets used for this project: rather than solely using past ‘Temperature’ values to predict future ‘Temperature’ values, one would also consider the past values of ‘Rainfall’ and ‘Sunshine’. The formulation of a $VARMA(p,d,q) \times (P,D,Q)_{12}$ model is as follows:

$$Y_t = \alpha + \sum_{i=1}^{p} \Phi_{i \times d} Y_{t-i \times d} + \sum_{i=1}^{P} \Phi_{i \times D \times 12} Y_{t-i \times D \times 12} + \varepsilon_t + \sum_{i=1}^{q} \Theta_{i \times d} \varepsilon_{t-i \times d} + \sum_{i=1}^{Q} \Theta_{i \times D \times 12} \varepsilon_{t-i \times D \times 12}$$

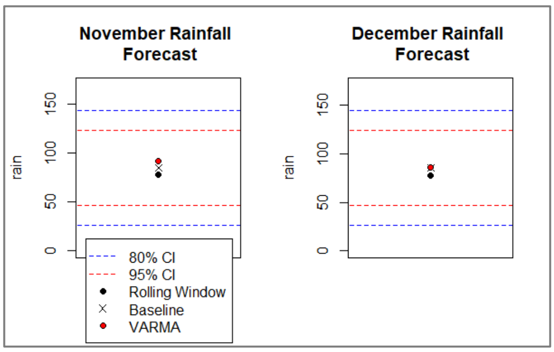
Where: Y_t is an $n \times 1$ matrix (n is the dimensions of the multivariate time series); α is an $n \times 1$ matrix of constants; Φ_j is an $n \times n$ matrix of autoregressive parameters at time $t - j$; ε_t is an $n \times 1$ matrix $\sim WN(0, \sigma_\varepsilon^2)$; Θ_j is an $n \times n$ matrix of moving average parameters at time $t - j$.

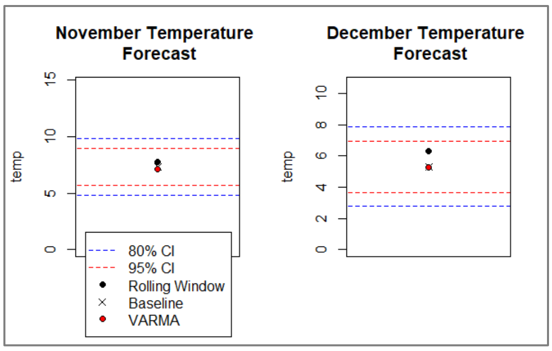
The 'MTS' package on R was used to fit VARMA models. There did not appear to be a function that allowed autofitting similar to **auto.arima()**. As such, multiple VARMA models were fitted with different parameters. Furthermore, I elected not to use a rolling window implementation partly due to the extended running time that such an implementation would have taken. Moreover, with a larger training dataset, one is more likely to fit a more optimised model. Therefore, I used the maximum data available to fit the various models. The input data itself had to be adjusted from the individual 'Temperature', 'Sunshine' and 'Rainfall' datasets into a multivariate series. Primarily, the individual sets are of different lengths, however, they are required to be of the same length. Therefore, the 'Temperature' and 'Rainfall' data sets were shortened to begin in 1919.

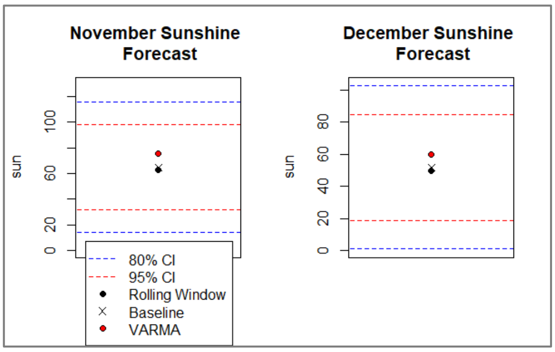
All the models were assumed to have a seasonality of period 12 with $D = 1$ and no trend $d = 0$ as these were characteristics of the most accurate ARIMA models fitted earlier. Below are the results of VARMA model fittings.

Model	**AIC**
$VARMA(1,0,1) \times (1,1,1)_{12}$	13.7731
$VARMA(1,0,0) \times (1,1,0)_{12}$	14.6914
$VARMA(1,0,0) \times (1,1,1)_{12}$	13.7637
$VARMA(1,0,0) \times (0,1,1)_{12}$	13.745
$VARMA(1,0,1) \times (0,1,1)_{12}$	13.7372
$VARMA(2,0,0) \times (0,1,1)_{12}$	13.7522
$\boldsymbol{VARMA(1,0,1) \times (0,1,2)_{12}}$	**13.6786**
$VARMA(1,0,1) \times (1,1,2)_{12}$	15.4284

$VARMA(1,0,1) \times (0,1,2)_{12}$ was chosen to predict the November and December values given that it had the lowest AIC value. The predictions of the VARMA model against those of the optimal rolling window models in question 3 and the base line predictions in question 2 are illustrated below.







The VARMA predictions are all well within a 95% confidence interval of the baseline predictions. Particularly, the VARMA 'Temperature' forecasts appear to be more aligned with the baseline forecasts than do the rolling window predictions. However, the rolling window forecasts are more in line with the baseline forecasts in regard to 'Sunshine'.

Appendix – R Script