

Buoy Project

Zijie Huang

2020/9/25

1. Abstract

This report seeks to use 20-years of data NDBC Station 44013 to check whether there is global warming. After data collection and cleaning, time series model and linear regression model were used to fit the data. Conclusions were drawn based on the results of models. Package Knitr (Yihui Xie, 2020) was used to compile this report.

2. Data Collection

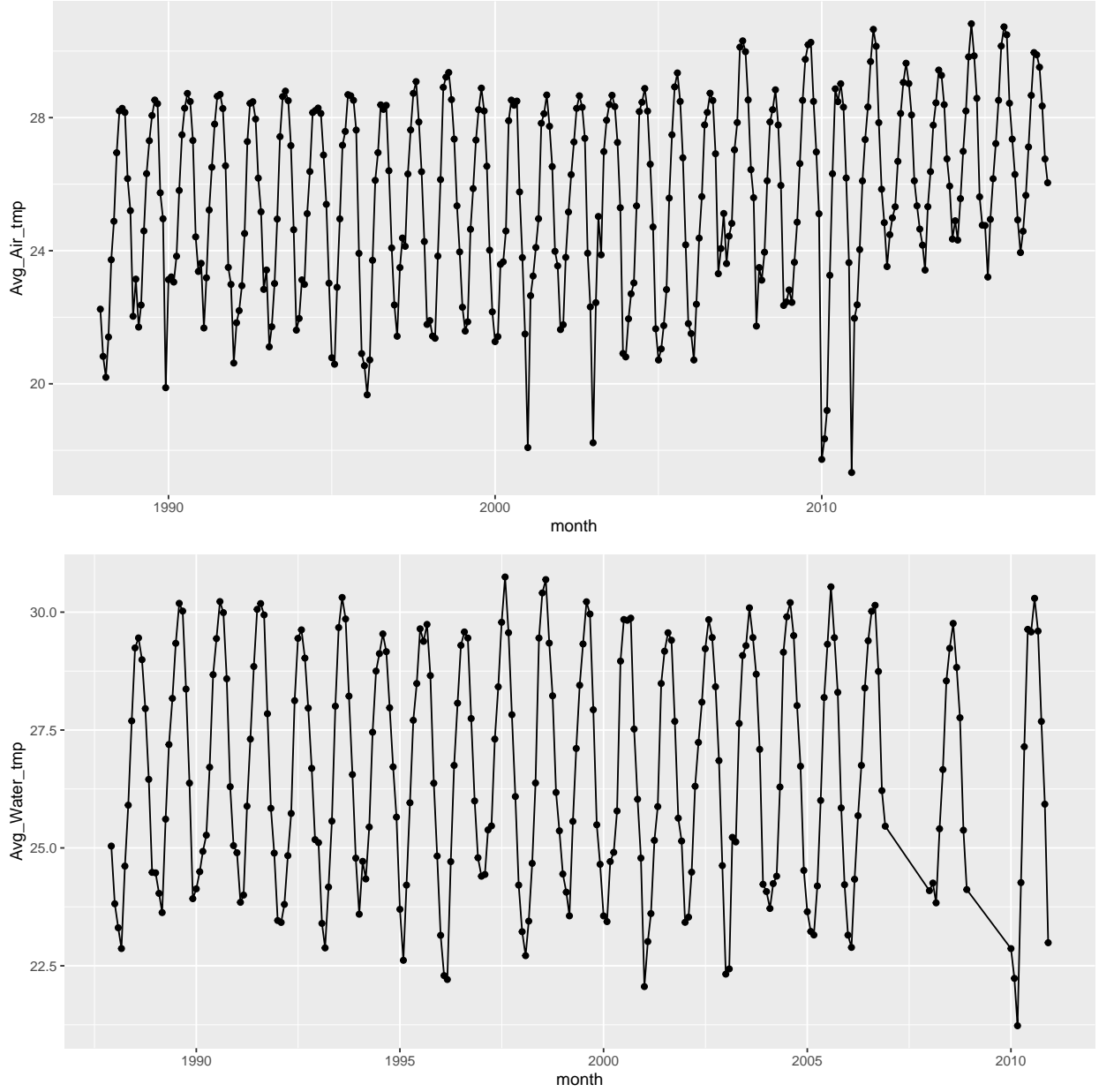
Data was abstracted from the NDBC website: https://www.ndbc.noaa.gov/station_history.php?station=44013. 30-years of data was collected from 1987 to 2016. Six variables were selected namely, year(YY), month(MM), day(DD), hour(hh), air temperature(ATMP) and sea surface temperature(WTMP). In order to see how the trend of temperature change, the only two related variables ATMP and WTMP were selected.

3. Data Cleaning

Package lubridate (Grolemund, Wickham 2011) and package tidyverse (Wickham et al., 2019) were used during data cleaning. Variable ATMP and WTMP were converted to *double* and a new variable 'YYYY_MM_DD' was added to the dataset. The new variable is *date* type. Data with temperatures higher than 90°C were deleted. The Average daily temperature was first calculated by taking the mean of 24 hours temperature. Then, the average monthly temperature was calculated by taking the mean of the average daily temperature per month. The reason why the temperature were calculated in this way was that all data was taken into consideration.

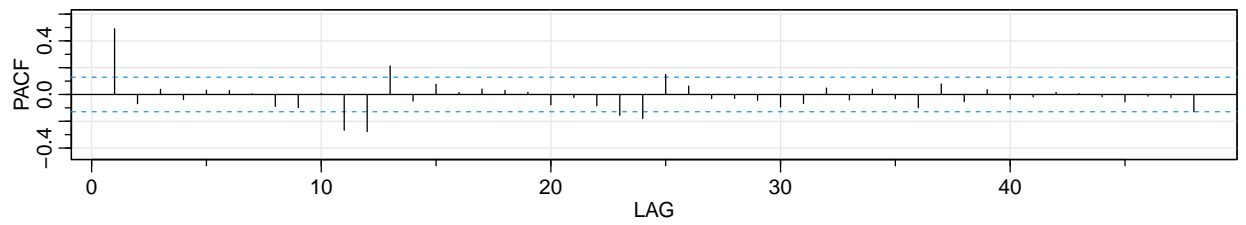
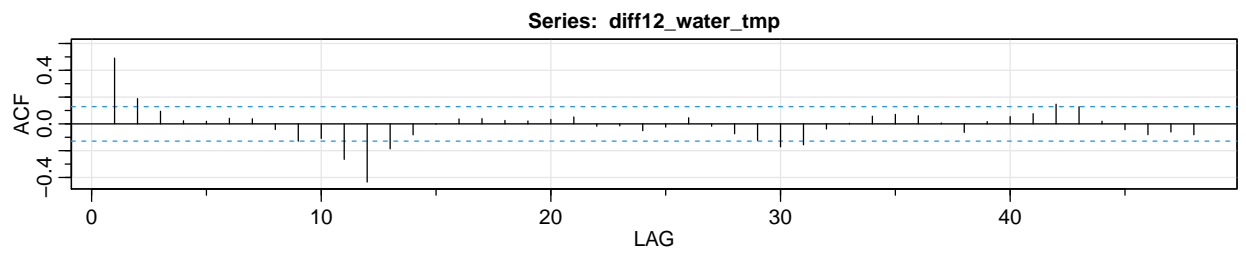
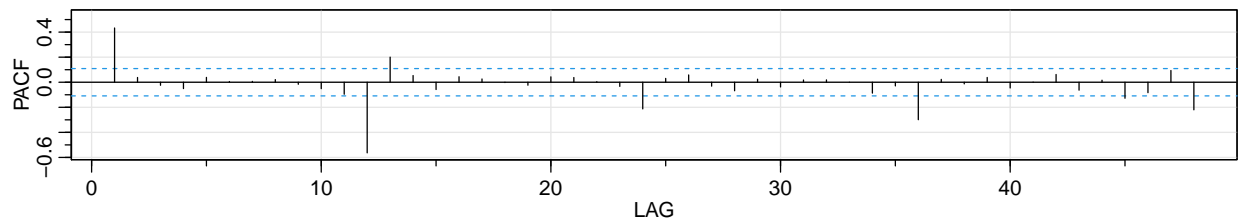
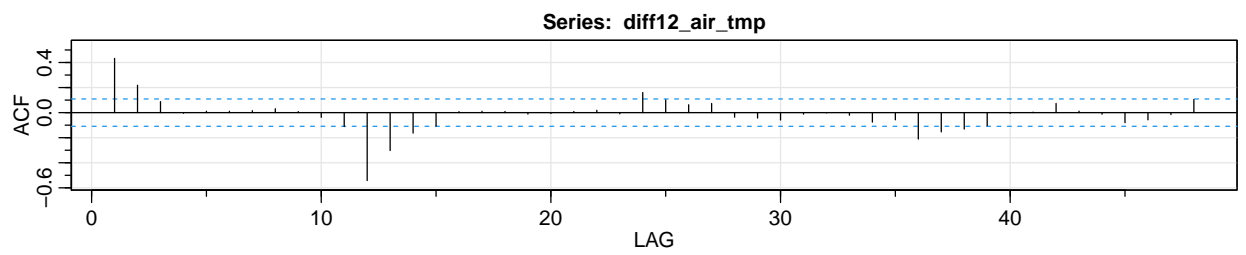
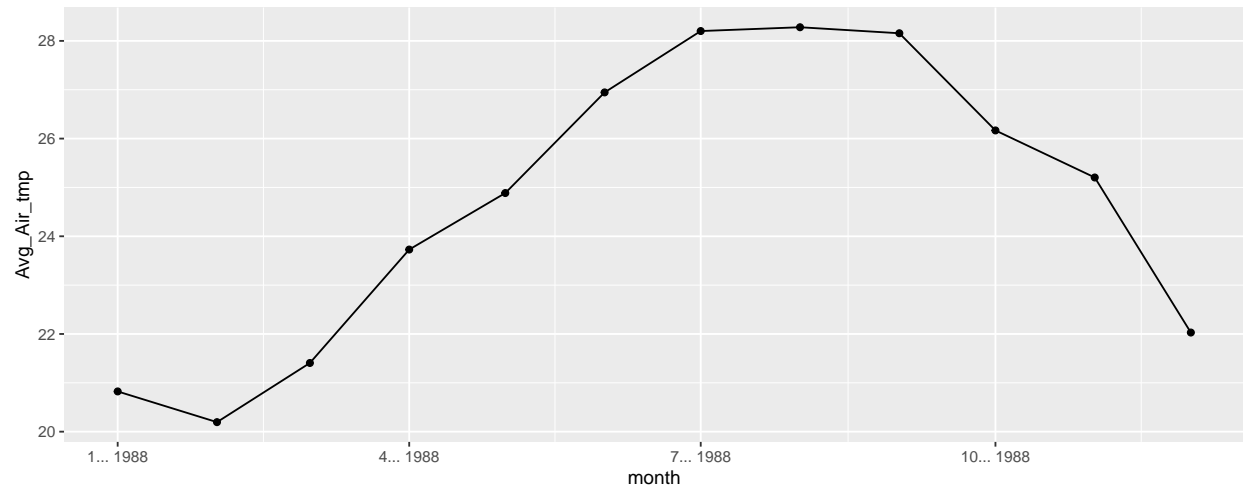
4. Model Fitting

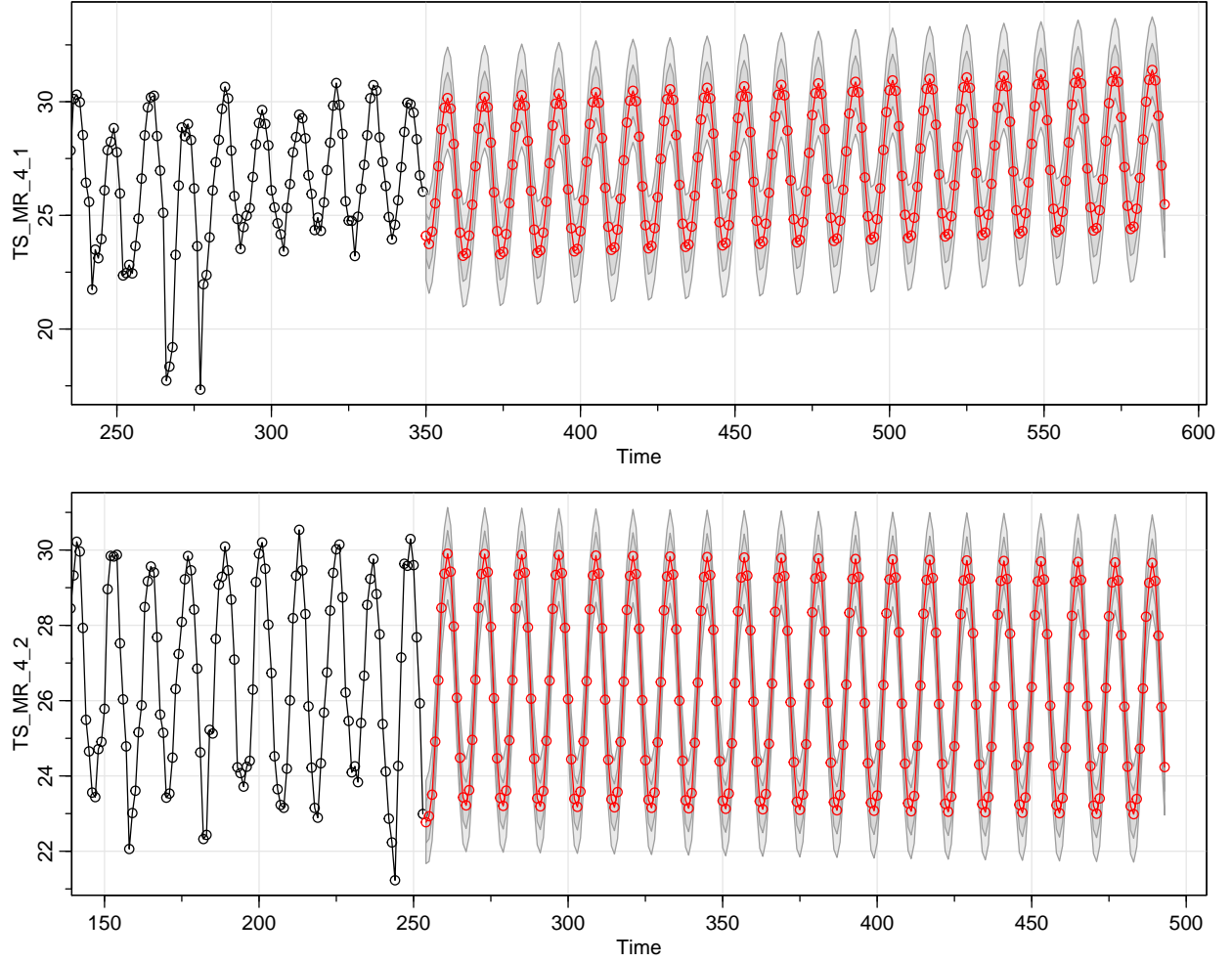
Package ggplot2 (Wickham, 2009) was used to plot data. Data was first plotted. The time series model was considered given the shape of the plot. The linear regression model was considered since it was a very simple model and the sign of the slope can be used to check the trend of climate change.



4.1 Time Series Model

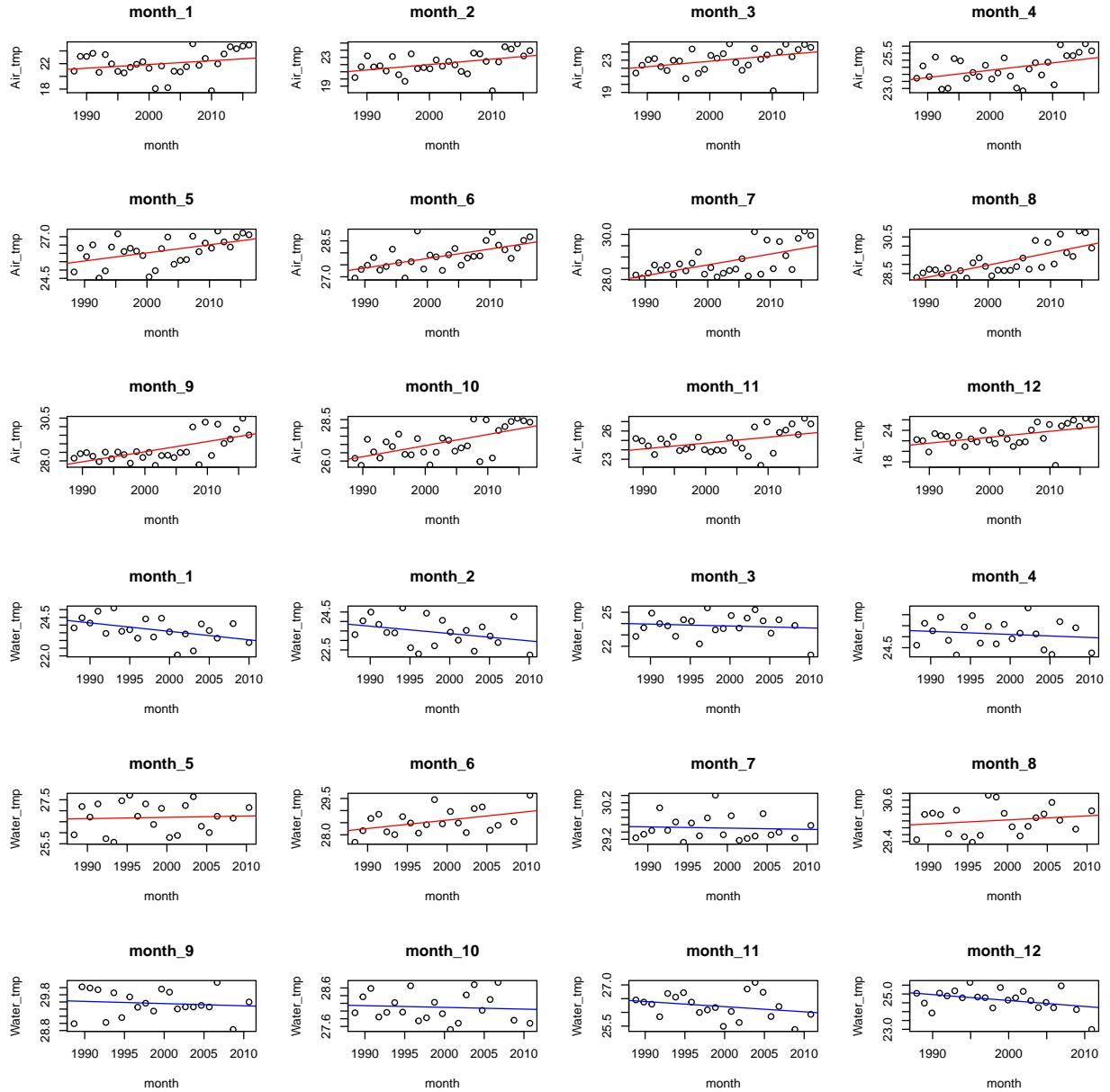
Package `astsa` (Stoffer, 2020) was used to perform time series analysis. Given it was a seasonal data, seasonal ARIMA model was considered. The trend of data within a year was checked. After setting the seasonal period s to be 12, ACF and PACF were plotted. By checking the plot, model $ARIMA(1, 0, 0) \times (0, 1, 1)_{12}$ was considered. The predicted data in the next 20 years was plotted. From this model, conclusion about global warming can't be drawn. No strong sign of rising or dropping in temperature was shown from the model. However, signs of a rise in air temperature and a drop in sea surface temperature were shown in the predicted data plot.





4.2 Linear Regression Model

Package `rstanarm` (Goodrich B, Gabry J, Ali I & Brilleman S., 2020) was used to perform linear regression analysis. For linear regression model, data was divided into 12 months. Data for each month was fitted to a linear regression model. All fitted lines and data were plotted on the same page. The red color indicated a upward trend and blue color indicated a downward trend. For air temperature, fitted lines from all months have a positive slope. For sea surface temperature, fitted lines from May to June and August have a positive slope. It is suggested that air temperature is increasing in these 30 years and sea surface temperature is increasing in the summer season.



5. Conclusion

From the time series model, both predicted dropping and rising trend of temperature were shown. For the linear regression model, a strong rising trend of air temperature was shown. Based on the data of air temperature, a sign of global warming was indicated. Based on the data of sea surface temperature, the sign of global warming is not that intense. However, no conclusion can be drawn regarding whether global warming caused the rising of air temperature in this area. Also, the time series model used in this report is rather simple. A more complex model is likely to give a more accurate prediction.

6. Reference

1. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
2. Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
3. Goodrich B, Gabry J, Ali I & Brilleman S. (2020). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.21.1 <https://mc-stan.org/rstanarm>.
4. Garrett Golemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software*, 40(3), 1-25. URL <http://www.jstatsoft.org/v40/i03/>.
5. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
6. David Stoffer (2020). astsa: Applied Statistical Time Series Analysis. R package version 1.10. <https://CRAN.R-project.org/package=astsa>
7. Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.29.

Yihui Xie (2015) *Dynamic Documents with R and knitr*. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963

Yihui Xie (2014) *knitr: A Comprehensive Tool for Reproducible Research in R*. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC. ISBN 978-1466561595