ARIMA Forecasting Exercise: Predicting the Global Sea Level Rise in the Short Run

Daria Tsvetkova Danil Islamov
Mikhail Martyanov
HSE University

February 20, 2022

ABSTRACT

TODO

Website: `https://ddtsvetkova.github.io/sealvl/`

1. Introduction

Combating the effects of climate change is one of the 17 Sustainable Development Goals put forward by the UN in 2015, and the rising seas is one of those effects. The root of such a change in sea levels is human-induced global warming, which causes the water to heat up and expand and the glaciers and ice sheets to melt more quickly. In essence, the rising global sea level is a time bomb with potentially devastating consequences, including the degradation of coastal ecosystems, increased intensity of tropical cyclones and storms, coastal erosion, salinization of aquifers, the flooding of coastal cities and, ultimately, imminent economic losses in property and infrastructure damages[3]. It is only natural that both country-level and international policymakers, business owners, and insurance companies should take an interest in a forecast of the future sea levels in order to take appropriate action. In this study we aim to solve such a short-run prediction problem using the tools of univariate time series linear modeling.

2. Data

The data used in this project come from the University of Hawaii Sea Level Center Fast Delivery gauges[1,2]. They are shown as change in sea level in millimeters compared to the 1993-2008 average (hence the below-zero values in Figure 1) and cover the years between 1970 and 2020 at a quarterly frequency, yielding a total of 203 data points. Note, though, that as our research question focuses on the future *rise* in sea levels, the point of reference (in this case, the 1993-2008 average) does not matter.

It is clear, both from general knowledge[3] and a glance at Figure 1, that the global sea level has been consistently rising over the years (albeit with some fluctuations), exhibiting an arguably linear trend. In just the last 50 years, it has increased by as much as 12 cm. Presumably, unless due measures are taken, the sea level will continue to rise. This fact alone is enough to reject the **stationarity** of the initial series, using essential judgement.

It might be prudent to digress and consider the possibility that the sea level is rising at an exponential, rather than linear, rate[4]: indeed, looking at a longer sample averaged from the two sources[1] (see Figure A.1 in Appendix A), one may find some (weak) evidence of an exponential trend. Still, we refrain from adopting this theory in practice due to our misgivings about combining different sources; besides, with our short forecast horizon, it might not make a difference.

Conditional on our conjecture that the trend is linear being true, we can test a hypothesis that the series is *integrated of order one (I(1))* by checking that its first difference is stationary. Neither essential nor visual analysis allow us to decidedly reject stationarity in this case (see Figure 2). The battery of formal tests likewise support stationarity, with the ADF and KPSS tests succeeding in and failing at rejecting their respective nulls (Table 1). Therefore, the data at hand are conceivably suitable for modeling with the ARIMA methodology.

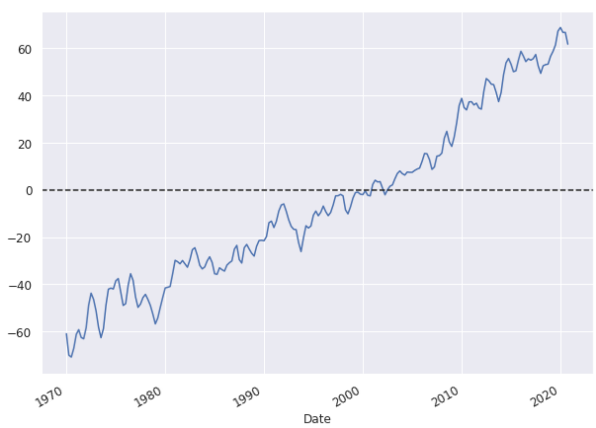


Figure 1: Global mean sea level rise, meas. relative to the 1993-2008 average sea level (mm)

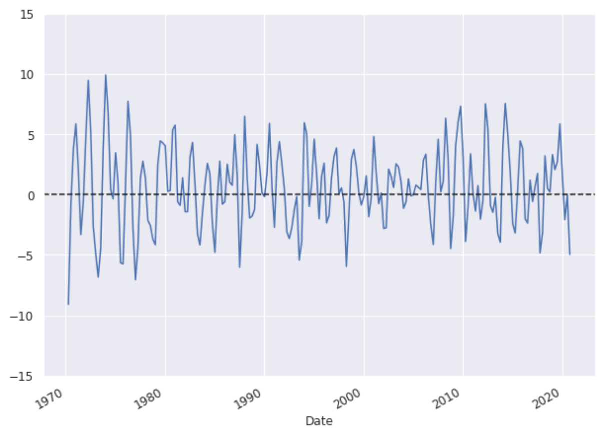


Figure 2: Global mean sea level rise, first difference (mm)

Table 1: Unit Root and Stationarity Tests Results on First Difference

	ADF Tests			*KPSS Tests*	
	H_0: Series has a unit root			H_0: Series is stationary	
	None	Intercept	Intercept and trend	Intercept	Intercept and trend
p-value	<0.01	<0.01	<0.01	>0.1	>0.1

Note: AIC is used to choose the number of lags for ADF tests. With SBC (not reported) the results are virtually the same. For exact critical and/or p-values see Appendix B.

3. Univariate linear model identification

Now we will turn to selecting the best ARIMA model to describe (and forecast) the global sea level rise. From now on we consider the *first difference* of the time series. Both its ACF and PACF (Figure 3) are significant only at the first 3 to 4 lags (with some lingering autocorrelation at higher lags). In other words, the textbook patterns of either pure AR (geometric decay in ACF, a cutoff in PACF) or MA (a cutoff in ACF, geometric decay in PACF) processes are absent. Perhaps some combination of AR and MA terms (i.e., an ARMA model) would explain this behavior.

The subsequent procedure is as follows. First, we estimate simple AR(p) and MA(q) models; then, looking at the residuals, we try experimenting with adding either MA or AR terms, increasing the complexity until the correlation has been removed and/or the AIC and SBC have decreased. For robustness, we also perform a simple *grid search* over p, q up to 10 seeking the combination that minimizes AIC and SBC (Tables A.3 and A.4, respectively).

Table A.2 reports selected results (note that ARIMA(p,0,q) here is equivalent to ARIMA(p,1,q) for the levels). A straight MA model initially performs better than a straight AR model in terms of both information criteria. Still, with the AR terms being highly significant, it might be advisable to add some.

The well-known heuristic of choosing the number of AR terms based on the number of significant PACF lags before the first insignificant one and the number of MA terms likewise based on ACF would suggest ARIMA(2,0,4) as our model of choice, and, incidentally, this model yields the lowest possible AIC (confirmed by the grid search).

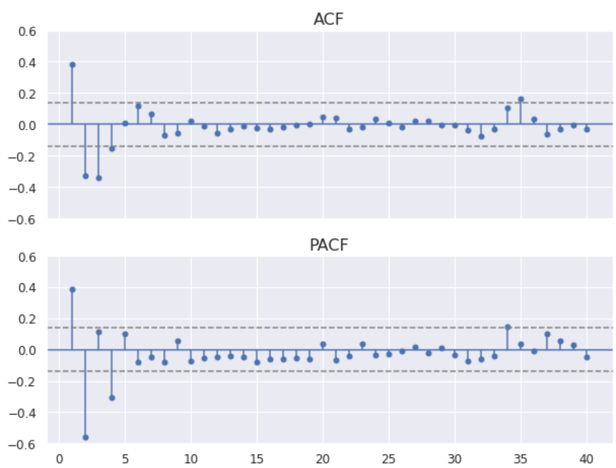


Figure 3: ACF and PACF of the first difference. The dashed lines are approximate 95% confidence intervals assuming that the true correlation is zero ($\pm 1.96/\sqrt{T}$)

Meanwhile, the lowest SBC (as well as the second-best AIC) is achieved by a more parsimonious ARIMA(1,0,3) (not surprisingly, given that SBC favours parsimony). Despite the fact that even ARIMA(0,0,4) does a fairly good job of generating the "white noise" residuals, we will mostly focus on the last two models for forecasting purposes, pitting the more complex (and thus more *overfit*-risky) model against the more parsimonious one.

References

Data from NOAA (2020)[2] via Our World in Data.[1] Out of those, only the series from the University of Hawaii Sea Level Center is used. For full description of the data see section 2.

[1] Ritchie, H., & Roser, M. (2020). CO_2 and Greenhouse Gas Emissions. Published online at OurWorldInData.org. Retrieved from: https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions. Dataset retrieved from: https://ourworldindata.org/grapher/sea-level-rise

[2] NOAA. (2020). Climate Change: Global Sea Level. NOAA Climate.gov. https://www.climate.gov/news-features/understanding-climate/climate-change-global-sea-level

[3] IPCC. (2019). Summary for Policymakers. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate [H.- O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, M. Nicolai, A. Okem, J. Petzold, B. Rama, N. Weyer (eds.)]. In press. https://www.ipcc.ch/srocc/chapter/summary-for-policymakers/

[4] Hansen, J., Sato, M., Hearty, P., Ruedy, R., Kelley, M., Masson-Delmotte, V., ... & Lo, K. W. (2016). Ice melt, sea level rise and superstorms: evidence from paleoclimate data, climate modeling, and modern observations that 2 C global warming could be dangerous. Atmospheric Chemistry and Physics, 16(6), 3761-3812. https://acp.copernicus.org/articles/16/3761/2016/acp-16-3761-2016.html

Appendix A

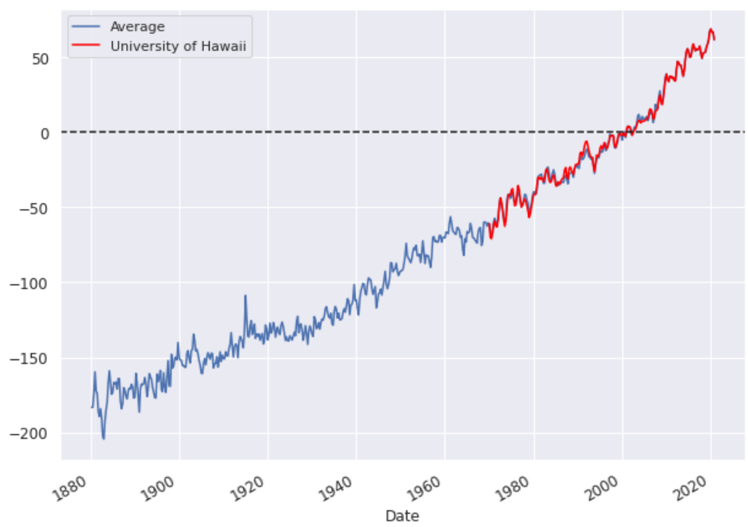


Figure A.1: Global mean sea level rise, meas. relative to the 1993-2008 average sea level (mm). The average is taken over the widely-cited Church & White dataset (1880-2009) and the University of Hawaii Sea Level Center (UHLSC) data (1970-2020)

Table A.2: Selected ARIMA Results

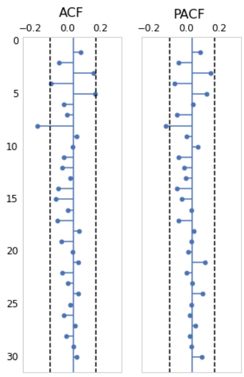
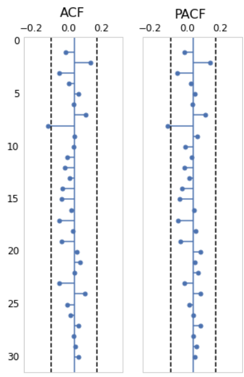
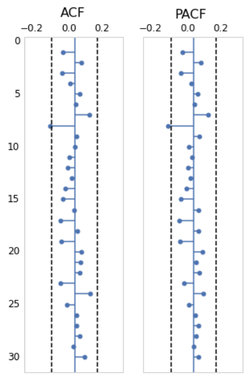
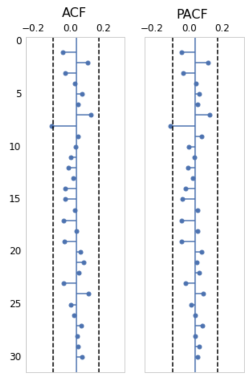
	Dependent variable: first difference			
	(4,0,0)	(0,0,4)	(2,0,4)	(1,0,3)
C	0.609***	0.606***	0.599***	0.605***
AR(1)	0.834***		1.276***	0.320***
AR(2)	-1.015***		-0.294**	
AR(3)	0.496***			
AR(4)	-0.422***			
MA(1)		1.418	0.181	1.124
MA(2)		-0.340	-1.812**	-0.732
MA(3)		-0.959	-0.180	-0.869
MA(4)		-0.193	0.813	
Log likelihood	-453.29	-404.65	-401.79	-403.81
AIC	918.58	821.30	819.57	819.62
SBC	938.46	841.18	846.08	839.50
Residuals				
	Here be forecast diagnostics			
Obs. (T)	203	203	203	203
Note:	*p<0.1; **p<0.05; ***p<0.01. For detailed output see Appendix B.			

Table A.3: Grid search results (AIC values)

	q										
p	0	1	2	3	4	5	6	7	8	9	10
0	1070.96	912.10	866.19	826.29	821.30	821.56	842.44	824.53	826.67	842.33	831.27
1	1038.80	907.75	900.74	819.62	823.47	820.13	821.49	822.93	824.01	825.96	827.96
2	956.57	867.46	868.52	828.96	819.57	830.82	820.77	821.84	823.52	823.87	825.79
3	952.64	869.27	861.39	827.69	821.34	820.87	827.07	822.26	826.73	828.26	827.80
4	918.58	846.18	827.11	828.25	840.57	835.34	823.51	827.76	825.41	825.25	828.97
5	906.54	847.64	846.86	825.03	824.17	830.25	824.48	827.86	834.71	830.96	833.43
6	896.54	847.78	846.51	850.07	826.13	828.47	830.27	828.40	829.34	832.76	835.34
7	897.37	849.11	848.27	853.20	850.75	829.84	833.43	829.62	830.75	831.56	834.64
8	891.22	841.45	846.98	852.54	850.07	829.46	841.06	828.93	832.68	836.28	832.80
9	887.47	841.28	843.78	825.71	846.90	829.71	833.11	833.05	835.98	837.31	838.43
10	880.59	840.80	841.02	842.53	849.55	829.82	829.47	834.89	834.10	842.86	838.01

Note: The smallest AIC is highlighted in red.

Table A.4: Grid search results (SBC values)

	q										
p	0	1	2	3	4	5	6	7	8	9	10
0	1077.59	922.04	879.44	842.85	841.18	844.76	868.94	854.35	859.81	878.77	871.02
1	1048.74	921.00	917.31	839.50	846.66	846.64	851.31	856.06	860.46	865.72	871.03
2	969.82	884.03	888.40	852.15	846.08	860.64	853.90	858.29	863.28	866.94	872.18
3	969.21	889.15	884.59	854.19	851.16	854.00	863.52	862.02	869.80	874.64	877.50
4	938.46	869.37	853.61	858.07	873.70	871.79	863.27	870.83	871.80	874.95	881.99
5	929.74	874.15	876.68	858.16	860.62	870.01	867.55	874.25	884.41	883.98	889.75
6	923.05	877.60	879.64	886.52	865.89	871.54	876.65	878.10	882.35	889.09	894.97
7	927.19	882.24	884.71	892.96	893.82	876.23	883.12	882.63	887.08	891.20	897.59
8	924.35	877.90	886.73	895.61	896.45	879.16	894.07	885.25	892.32	899.23	899.07
9	923.91	881.04	886.85	872.10	896.60	882.72	889.43	892.69	898.93	903.57	908.00
10	920.35	883.87	887.41	892.23	902.56	886.15	889.11	897.84	900.36	912.44	910.90

Note: The smallest SBC is highlighted in blue.

Appendix B

Relevant Python code is available here: `https://ddtsvetkova.github.io/sealvl/code.html`