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

Daria Tsvetkova Danil Islamov
Mikhail Martyanov
HSE University

February 28, 2022

ABSTRACT

Rising sea levels is a threat not to be taken lightly. The present study aims to solve a short-run prediction problem by selecting an appropriate univariate linear model to forecast the global sea level rise. Using quarterly data for the years 1970-2020, we suggest a suitable stationary transformation and evaluate the quality of both fit and prediction power of 121 models. Based on this analysis, the resulting forecast is provided.

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 supports 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.

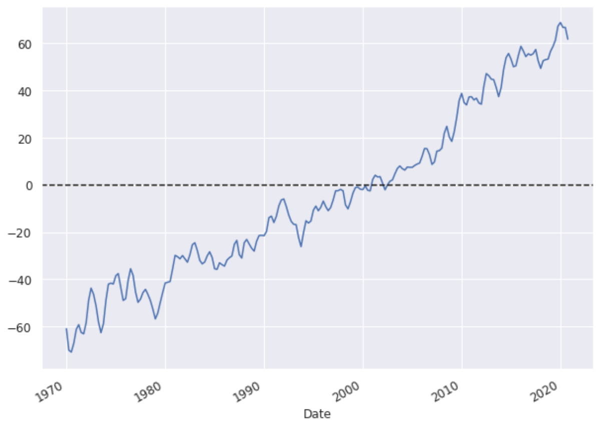


Figure 1: Global mean sea level rise, meas. relative to the 1993-2008 average sea level (mm)

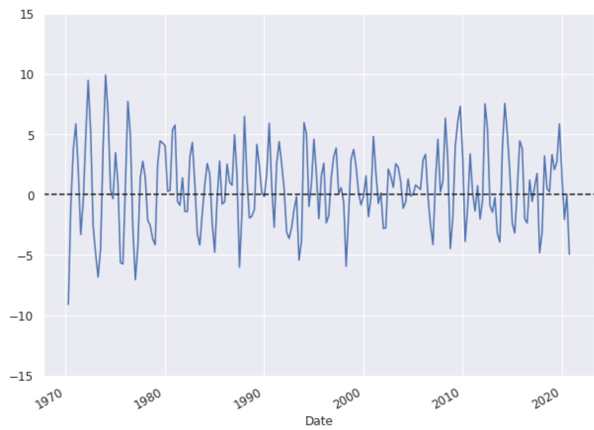


Figure 2: Global mean sea level rise, first difference (mm)

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. During this stage 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 second-best AIC (confirmed by the grid search).

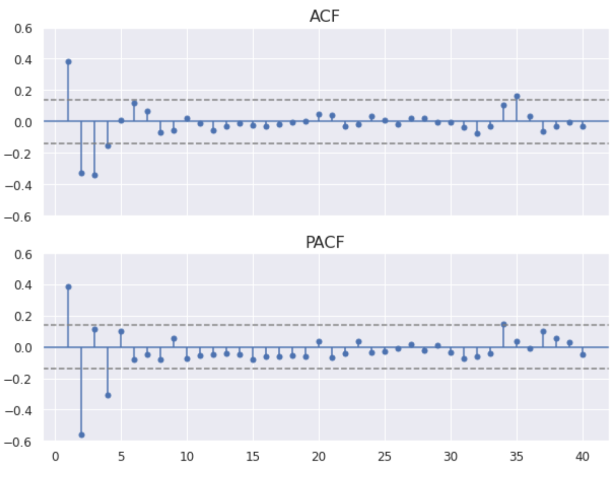


Figure 3: ACF and PACF of the first difference. The dashed lines are approximate 95% confidence intervals assuming that the true correlation is zero ($\pm 1.96/\sqrt{T}$)

At the same time, both the lowest SBC and the lowest AIC are achieved by a more parsimonious ARIMA(1,0,3). ARIMA(1,0,5) is also a valid contender, with its third-best AIC and one of the lower SBCs. All of the competing models do a fairly good job of generating the "white noise" residuals, as evidenced by the results of the Ljung-Box Q tests (found in Appendix B). However, the best results of the in-sample diagnostics do not guarantee the superior quality of out-of-sample forecasts, which we will deal with in the next section.

4. Prediction

Indeed, the last stage's leader in terms of fit, ARIMA(1,1,3)[†], fails miserably when it comes to prediction. This fact is confirmed by the following procedure. First, we divide the sample into training, validation, and testing sets (Figure A.2). Rather conservatively, we allocate 80% of our data to the former and the most recent 20% to the latter two (making them 5 years long each). Then, for robustness' sake, we repeat the *grid search* over ABC and SBC on the training subsample, with approximately the same results (not reported here; see Appendix B). A separate search over the forecast diagnostics on the validation set, however, uncovers models that, while falling behind at the training stage, far outpace the others in forecast quality. They are reported in Table A.5 and Figure A.3, with ARIMA(1,1,3) for contrast. Table A.6 presents the results of the grid search over RMSE, to emphasize the relative sizes of that statistic with different ARIMA hyperparameters. Crucially, ARIMA(1,1,3) underestimates the global sea level rise (Figure A.3), which in practice would pose a threat of underpreparedness to the regions abiding by that model.

What might be the reason of such a sudden change in the ranks of the best-performing models? A fairly accessible explanation for the inferior prediction quality of the last stage's leaders is *overfitting*. The fact that they better perform on the training set than they do on the testing one is a sign of their poor ability to generalize to "unseen" data. It might be that these models are picking up random fluctuations from the training data.

As the key goal of this study is prediction, the forecast accuracy is paramount. This leads us to essentially throw the "best fit" models out the window. Instead, we use ARIMA(9,1,0), the leading one according to several forecast assessment measures, to predict sea level rise over the testing period (Figure 4). Fortunately, the "good" forecast diagnostics do not deteriorate completely as we move from the validation to the testing sample, with MA%E even marginally improving. Moreover, the observed sea level rise between 2016 and 2020 falls well within the 95% confidence interval of our forecast.

[†] Now we return to considering *levels* to make it easier to look at forecast diagnostics.

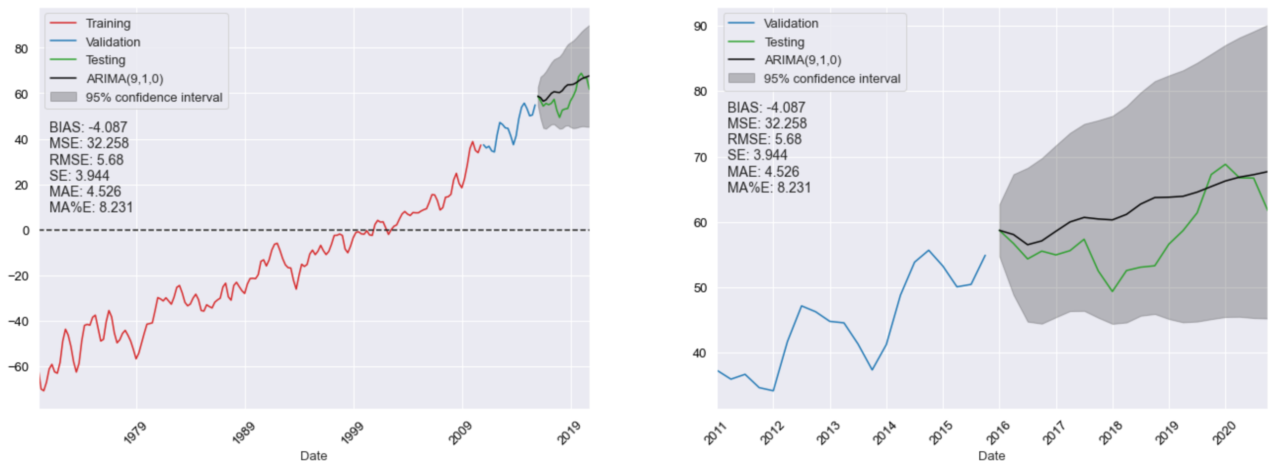


Figure 4: Forecast on the testing set

To conclude, the tools of univariate time series linear modeling (the ARIMA methodology) are reasonably adequate to the forecasting problem at hand: namely, predicting the global sea level rise in the short run. As expected based on essential judgement, the sea levels are predicted to grow. The soundness of this forecast was ensured by the *grid search* over the model hyperparameters and verified by both visual and technical analysis. The results of the study at hand may provide food for thought for different kinds of economic agents.

5. Teamwork evaluation

We deem everyone's participation and contribution fair, commensurate, and satisfactory. We thank each member of our team for their assistance and feedback during our collaboration on this project.

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

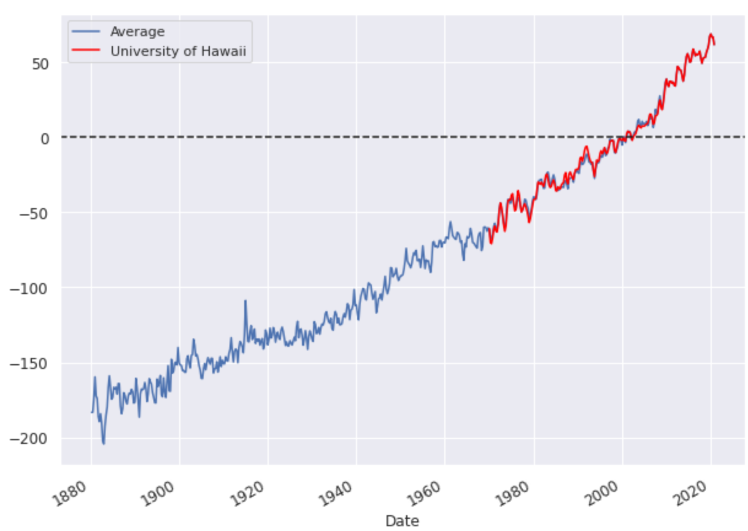
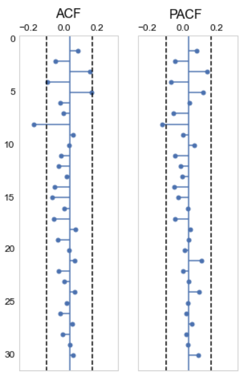
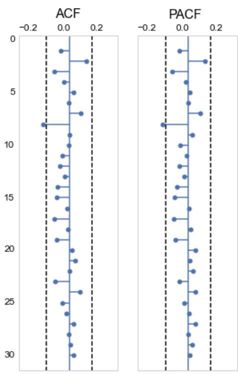
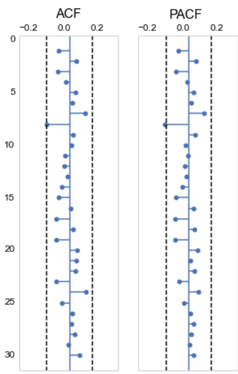
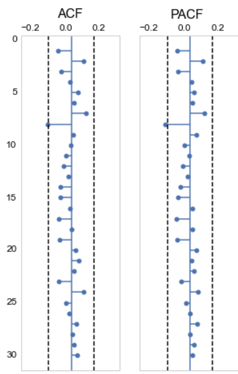
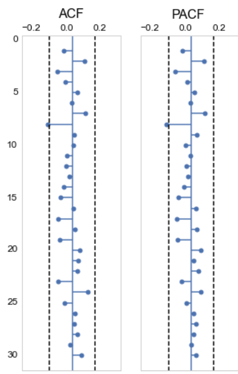


Figure A.1: Global mean sea level rise, meas. relative to the 1993-2008 average sea level (mm). The average is taken over the widely-cited Church & White dataset (1880-2009) and the University of Hawaii Sea Level Center (UHLSC) data (1970-2020)

Table A.2: Selected ARIMA Results

	Dependent variable: first difference				
	(4,0,0)	(0,0,4)	(2,0,4)	(1,0,3)	(1,0,5)
C	0.609***	0.606***	0.605***	0.605***	0.599***
AR(1)	0.834***		1.105	0.320***	0.965***
AR(2)	-1.015***		-0.232		
AR(3)	0.496***				
AR(4)	-0.422***				
MA(1)		1.415***	0.350	1.123***	0.472
MA(2)		-0.340**	-1.625	-0.732***	-1.705***
MA(3)		-0.958***	-0.307	-0.868***	-0.656
MA(4)		-0.192**	0.670		0.707**
MA(4)					0.185
Log likelihood	-453.29	-404.65	-401.95	-403.81	-402.09
AIC	918.58	821.30	819.89	819.62	820.17
SBC	938.46	841.18	846.40	839.50	846.68
Residuals					
Obs. (T)	203	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)

p	q: 0	1	2	3	4	5	6	7	8	9	10
0	1070.96	912.10	866.82	826.29	821.30	821.57	822.95	824.53	826.53	828.44	821.04
1	1038.80	907.75	903.68	819.62	823.48	820.17	821.55	823.02	824.05	FC	828.02
2	956.57	867.46	868.52	828.99	819.89	824.61	824.18	824.80	825.62	826.87	829.32
3	952.64	869.27	861.39	827.67	821.35	822.51	823.39	827.82	827.02	828.96	828.46
4	918.58	846.18	827.81	828.30	839.10	834.39	823.81	826.99	824.90	823.86	831.56
5	906.54	847.64	846.86	824.97	824.13	830.96	829.15	828.66	828.48	834.20	830.56
6	896.54	847.78	846.51	824.62	824.71	829.81	826.57	827.66	833.50	832.30	830.39
7	897.37	849.11	848.27	851.83	850.45	834.44	827.63	834.53	831.80	834.13	832.92
8	891.22	841.45	846.98	825.35	824.05	833.35	830.80	834.20	828.69	834.35	834.38
9	887.47	841.28	843.76	825.66	827.05	828.39	828.83	832.67	835.14	828.77	834.51
10	880.59	840.80	841.02	842.29	828.63	832.03	829.54	836.09	828.82	833.78	840.66

Note: The smallest AIC is highlighted in red. FC stands for "failed to converge in 80 iterations".

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

p	q: 0	1	2	3	4	5	6	7	8	9	10
0	1077.59	922.04	880.07	842.86	841.18	844.76	849.45	854.35	859.66	864.89	860.80
1	1048.74	921.00	920.25	839.50	846.67	846.68	851.37	856.15	860.49	FC	871.09
2	969.82	884.03	888.40	852.18	846.40	854.43	857.31	861.24	865.38	869.94	875.71
3	969.21	889.15	884.59	854.18	851.17	855.64	859.83	867.57	870.10	875.35	878.16
4	938.46	869.37	854.32	858.12	872.23	870.83	863.57	870.06	871.28	873.56	884.57
5	929.74	874.15	876.68	858.10	860.58	870.72	872.22	875.04	878.18	887.21	886.88
6	923.05	877.60	879.64	861.07	864.47	872.88	872.96	877.36	886.51	888.63	890.03
7	927.19	882.24	884.71	891.59	893.52	880.82	877.33	887.54	888.12	893.77	895.87
8	924.35	877.90	886.73	868.43	870.43	883.05	883.81	890.53	888.33	897.30	900.64
9	923.91	881.04	886.84	872.04	876.75	881.40	885.15	892.31	898.09	895.03	904.09
10	920.35	883.87	887.41	891.98	881.64	888.36	889.18	899.04	895.08	903.36	913.55

Note: The smallest SBC is highlighted in blue. FC stands for "failed to converge in 80 iterations".

Table A.5: Selected Forecast Diagnostics

	Dependent variable: sea level rise (levels)				
	(1,1,3)	(9,1,0)	(7,1,1)	(5,1,1)	(2,1,5)
BIAS	14.758	0.698	0.336	0.736	1.768
MSE	259.733	19.192	20.466	20.468	23.778
RMSE	16.116	4.381	4.524	4.524	4.876
SE	6.475	4.325	4.511	4.464	4.545
MAE	14.758	3.763	3.886	3.943	4.119
MA%E	31.787	8.491	8.937	8.927	8.967
Obs. (Validation set)	20	20	20	20	20

Note: For full *grid search* results and detailed output see Appendix B.

Table A.6: Grid search results (RMSE on the validation set)

	q										
p	0	1	2	3	4	5	6	7	8	9	10
0	FLB	FLB	FLB	4.91	5.88	10.07	13.71	14.82	FC	15.87	15.28
1	FLB	FLB	FLB	16.12	15.57	14.83	14.57	14.49	14.34	FC	FC
2	FLB	FLB	4.60	15.83	14.65	4.88	6.98	16.85	16.73	16.07	15.92
3	FLB	FLB	14.36	15.76	15.79	16.68	15.12	15.11	15.95	14.45	14.40
4	FLB	4.56	15.35	15.28	14.13	16.46	14.94	4.77	14.95	8.47	16.16
5	FLB	4.52	12.65	13.85	14.83	15.98	17.21	15.86	15.48	16.18	14.72
6	FLB	4.66	14.73	14.03	13.79	15.46	14.02	15.53	14.59	12.76	13.18
7	4.52	4.52	13.97	11.13	13.36	13.72	10.16	20.35	14.34	19.97	13.76
8	5.03	5.01	16.63	15.46	15.41	11.37	19.62	16.47	17.56	15.69	15.51
9	4.38	4.75	14.96	11.17	11.37	17.90	13.78	13.91	16.79	15.58	17.16
10	5.69	5.13	17.52	16.04	11.72	11.93	11.72	16.69	10.83	12.07	9.79

Note: The smallest RMSE is highlighted in orange. FC stands for "failed to converge in 80 iterations". Here, the models failing the Ljung-Box Q tests (at least 1 significant correlation at lags up to 30 remains) are explicitly marked as FLB. For results on other forecast diagnostics see Appendix B.

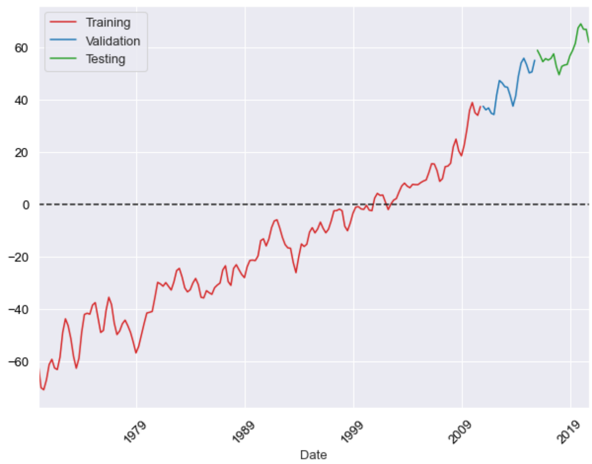


Figure A.2: Division into training (80%), validation, and testing sets

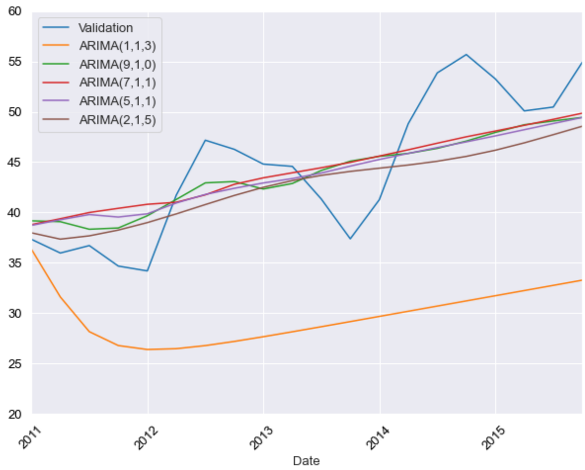


Figure A.3: Performance of selected models on the validation set (see Table A.5 for details)

Appendix B

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

NB: Most, if not all, results were cross-checked using EViews.