Time Series ARIMA Study of Antarctic Glacier Melting

Mason Chen
Stanford Online High School
San Jose, CA 95131
Mason.chen.training@gmail.com

Abstract

This STEM paper would study the Time Series Antarctic Glacier Mass from 2002 Apr. to 2021 Mar. The objective of this paper is to forecast the Antarctic Glacier Mass level for 2021-2041. Among four STEM components: Science is Geoscience of the Glacier; Technology is using the GRACE-FO satellites to collect Glacier Ice Sheet Mass data; Engineering would focus on COVID-19 factor on the Glacier melting rate, and Mathematics is mainly on Time Series ARIMA models. Both non-Seasonal and Seasonal ARIMA models were studied and compared. Both the 12-month seasonal pattern and long-term year to year trend were significantly observed. The Glacier meting rate was becoming 2% faster based on the Seasonal ARIMA model. Smoothing models were also significantly identified in the Seasonal ARIMA model to smooth out the random noise component to enhance the Time Series Trend and Seasonal component to enhance the forecasting model. Forecasting Glacier Melting for 2021-2041 would be a challenging task to address both seasonal and trend components for a longer horizontal time from today. The prediction interval would become too wide to predict the future Glacier melting rate if more than 5 years away. Seasonal ARIMA model could provide a better fit than the non-seasonal ARIMA model, STEM methodology is a powerful and holistic way for conducting Scientific research project by modern GRACE-FO Technology in a practical Engineering sense through a Mathematical ARIMA Forecasting analysis.

Keywords
ARIMA, Time Series Forecast, Geoscience, Antarctic Glacier, GRACE-FO

1. Introduction
This project would study the Antarctic Glacier Mass data from 2002-2021 March. Objective: Use the Time Series ARIMA platform to examine the time series Glacier data to predict the Glacier crisis for the next twenty years (2021-2041).

1.1 STEM Methodology
Science: Geoscience (Earth Science) is the study of Earth. Geoscience includes so much more than rocks and volcanoes, it studies the processes that form and shape Earth's surface, the natural resources we use, and how water and ecosystems are interconnected. Geoscience uses tools and techniques from other science fields as well, such as chemistry, physics, biology, and math. Technology: The Gravity Recovery and Climate Experiment Follow-On (GRACE-FO) Satellites. Engineering: study the Antarctic Glacier Melting Crisis for 2021-2041. Mathematics: Statistical Time Series ARIMA Modeling analysis of the Glacier Ice Sheet Mass data.

1.2 Scientific Research Literature and Technology: GRACE-FO
The climate has been becoming out of control due to the Global Warming effect Arnold (2011). The Gravity Recovery and Climate Experiment Follow-On (GRACE-FO) mission is a partnership between NASA and the German Research Centre for Geosciences (GFZ). GRACE-FO is a successor to the original GRACE mission, which orbited Earth from 2002-2017. GRACE-FO will carry on the extremely successful work of its predecessor while testing a new technology designed to dramatically improve the already remarkable precision of its measurement system. Global surface mass anomalies observed by the GRACE-FO satellites (for the month indicated on the map). Over land, red colors indicate below-average terrestrial water amounts, while blue colors show above-average water amounts (including ice, snow, soil moisture and groundwater. Over oceans, red colors indicate below-average ocean bottom pressure, while blue colors show above-average bottom pressure. Ocean bottom pressure changes are related to large-scale ocean current variations, as well as overall sea level changes from ocean mass changes.

1.3 Engineering: Antarctic Glacier Melting Crisis

An Antarctic glacier larger than the UK is at risk of breaking up after scientists discovered more warm water flowing underneath it than previously thought. Over the past few years, teams of scientists have been crisscrossing the remote and inaccessible region on Antarctica's western edge to try to understand how fast the ice is melting and what the consequences for the rest of the world might be. "What happens in west Antarctica is of great societal importance," said Dr Robert Larter, a scientist with the British Antarctic Survey and principal investigator with the International Thwaites Glacier Collaboration, the most ambitious research project ever carried out in Antarctica. This is the biggest uncertainty in future sea level rise.

1.4 Mathematics: Time Series ARIMA and Forecast

Time Series Analysis and Forecasting modeling were utilized on the GRACE-FO Glacier Mass data. Climatology research has used Time Series and Forecasting model such as ARIMA to forecast the weather temperature to study the global warming trend Baillie (2002), Bindoff (2013). In this paper, we would compare different Time Series ARIMA Models on the Forecasting Capability for next 20 years (2021-2041).

2. Data Collection and Sampling Plan

2.1 GRACE-FO Antarctic Glacier Data and Sampling Plan

The data source for this paper is from the NASA GRACE-FO satellites' data of the Antarctic Ice Sheet Mass Trend as shown in Figure 1. The sampling plan is based on the monthly average of the images collected from satellites.

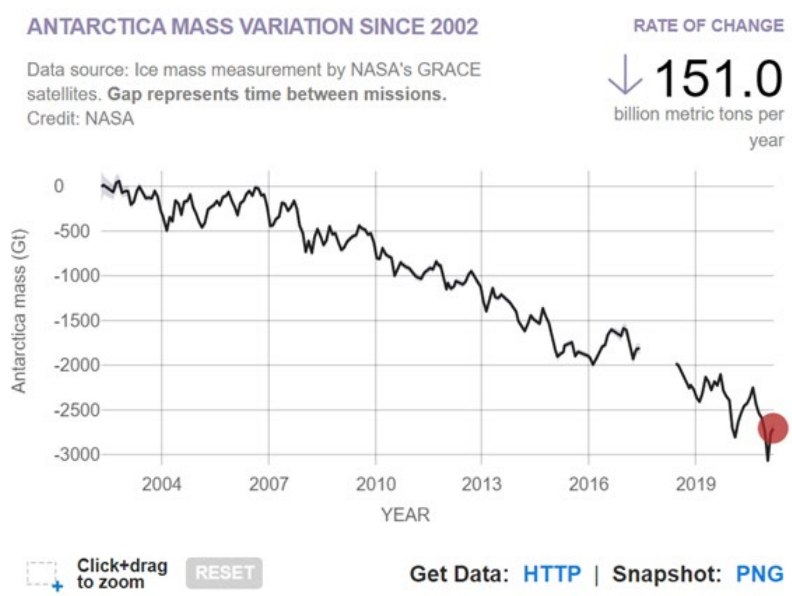

Figure 1. Antarctic Monthly Mass Trend

The Glacier Mass raw data was uploaded to the JMP platform from the NASA GRACE-FO website as shown in Figure 2.

2002
Source
Columns (6/0)
Antarcti...tonnes)
Year
Month
Year-Month
Antarcti...rended)
Antarcti...it cycle)
Rows
All rows 219
Selected 0
Excluded 0
Hidden 0
Labeled 0

	Antarctic mass (Gigatonnes)	Year	Month	Year-Month	Antarctic mass (Gigatonnes) (Detrended)	Antarctic mass (Gigatonnes) (Remove 12 unit cycle)
1	•	2002	1	01/2002	•	•
2	•	2002	2	02/2002	•	•
3	•	2002	3	03/2002	•	•
4	0	2002	4	04/2002	-312.2540246	141.67839272
5	18.36	2002	5	05/2002	-281.1661726	51.559262294
6	•	2002	6	06/2002	•	•
7	•	2002	7	07/2002	•	•
8	-59.82	2002	8	08/2002	-321.1626167	-285.6739764
9	45.54	2002	9	09/2002	-203.0747647	-166.6549122
10	62.69	2002	10	10/2002	-173.1969127	-78.98839272
11	-69.03	2002	11	11/2002	-292.1890607	-102.2292623
12	-49.78	2002	12	12/2002	-260.2112087	34.395583658
13	-48.71	2003	1	01/2003	-246.4133567	130.28564995
14	-200.03	2003	2	02/2003	-385.0055048	25.823976382
15	-171.49	2003	3	03/2003	-343.7376528	40.70491224
16	-43.66	2003	4	04/2003	-203.1798008	98.018392724
17	0.79	2003	5	05/2003	-146.0019488	33.989262294
18	•	2003	6	06/2003	•	•
19	-128.94	2003	7	07/2003	-250.2762448	-307.9356499
20	-122.41	2003	8	08/2003	-231.0183929	-348.2639764
21	-130.92	2003	9	09/2003	-226.8005409	-343.1149122
22	-48.06	2003	10	10/2003	-131.2126889	-189.7383927
23	-107.58	2003	11	11/2003	-178.0048369	-140.7792623
24	-273.11	2003	12	12/2003	-330.8069849	-188.9344163

Figure 2. Glacier Mass Monthly Row Data File

3. Time Series Non-Seasonal ARIMA Model

Conduct JMP 16 Time Series Non-Seasonal ARIMA on the Glacier Mass data. There are two objectives of this Section 3: (1) Can Non-Seasonal ARIMA Forecasting address both Glacier Mass Trend and Seasonal components? (2) Which is the Optimal Non-Seasonal ARIMA model for Antarctic Glacier Mass forecasting?

3.1 Non-Season ARIMA Algorithm

ARIMA has three mathematical components box (1994): Autoregression (AR), Integration (I) and Moving Average (MA). Autoregression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values. Integrated (I): represents the differencing of raw observations to allow for the time series to become stationary, i.e., data values are replaced by the difference between the data values and the previous values. Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations. There are two ARIMA types: one is non-Seasonal and one is Seasonal. The difference is whether there is a fixed Seasonal component observed or detected in the Time Series Data. A Non-Seasonal ARIMA model is commonly denoted ARIMA(p,d,q). The AR "p" number considers the Autoregression AR module by integrating the historical values in exponentially decaying algorithm. The I "d" number considers the Integration Differencing I module by differencing the data points to detect the trend component. The MA "q" number considers the Moving Average MA module by smoothing the error term exponentially. If any of p, d, or q are zero, the corresponding letters are often dropped. For example, if p and dare zero,
then the model would simply be a moving average model, denoted as MA(q). The Seasonal ARIMA model would be addressed in Section 4.

3.2 Non-Seasonal ARIMA Analysis

JMP Non-Seasonal ARIMA platform was conducted as shown in Figure 3 across all potential ARIMA models up to level one for each three ARIMA components. The first focus of the Basic Time Series Analysis is to detect any Seasonal component at any fixed frequency (lag= 12). There was not clear seasonal pattern at lag=12 from Autocorrelation, Partial Correlation, Variogram and AR coefficient in Figure 4. There are two scenarios: one is no true seasonal component and the other one is the seasonal component may be masked by the stronger trending component. The models were ranked based on the default AIC criteria. Ranking was quietly consistently among all goodness fit index Burnham (2004, 2011). The tip two models are ARIMA (1,1,1) or I (1).

Model Comparison

Report	Graph	Model	DF	Variance	AIC ^	SBC	RSquare	-2LogLH	Weights	.2 .4 .6 .8	MAP
☑	☑	ARIMA(1, 1, 1)	175	10197.384	2151.3981	2160.9435	0.986	2145.3981	0.963334		74.959
☑	☑	I(1)	177	10785.478	2159.0394	2162.2212	0.986	2157.0394	0.021110		62.700
☐	☐	IMA(1, 1)	176	10846.548	2161.0360	2167.3996	0.986	2157.036	0.007779		63.129
☐	☐	ARI(1, 1)	176	10846.587	2161.0366	2167.4002	0.986	2157.0366	0.007777		63.049
☐	☐	AR(1)	193	11713.654	2387.0054	2393.5514	0.961	2383.0054	0.000000		
☐	☐	ARMA(1, 1)	192	11774.366	2389.0023	2398.8213	0.961	2383.0023	0.000000		
☐	☐	MA(1)	193	239036.1	2971.5227	2978.0687	0.647	2967.5227	0.000000		
☐	☐	ARIMA(0, 0, 0)	194	685348.46	3174.7316	3178.0046	0.000	3172.7316	0.000000		

Figure 3. Compare various Non-Seasonal ARIMA models

3.3 Interpret ARIMA (0,1,0) Model with Constant c

There are two major findings in Figure 3: (1) Integration component is 1 for both top two models, (2) AR and MA components are necessary? Here in Section 3.3, the first finding is related to ARIMA (0,1,0) or I (1) model. Different ARIMA (0,d,0) with constant c models were listed in Figure Fig.4. In this paper, Integration component was limited to d= (0,1) range. In Figure 3, d=1 was detected over d=0 for Forecasting the Antarctic Glacier Mass data.

- If $c = 0$ and $d = 0$, the long-term forecasts will go to zero.
- If $c = 0$ and $d = 1$, the long-term forecasts will go to a non-zero constant.
- If $c = 0$ and $d = 2$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 0$, the long-term forecasts will go to the mean of the data.
- If $c \neq 0$ and $d = 1$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 2$, the long-term forecasts will follow a quadratic trend.

Figure 4. ARIMA (0,d,0) models

Whether the constant c should be set zero or not would impact the Forecasting Trend component: a non-zero constant or a straight line. In Figure 5, the Forecasting for the top two ARIMA models has shown a straight line which has indicated the constant c should not be zero.

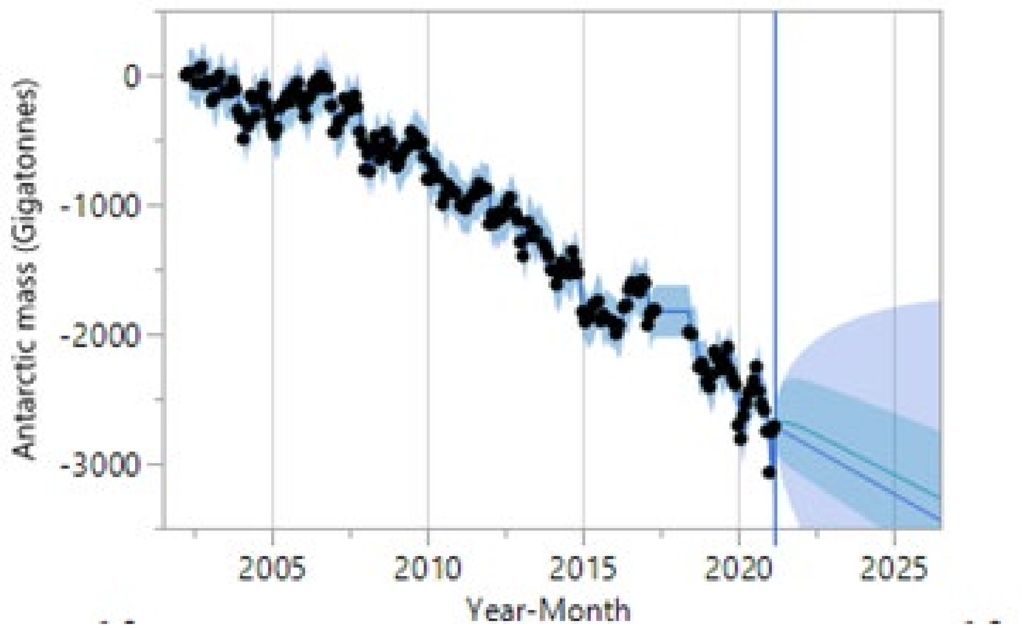

Figure 5. Non-Seasonal ARIMA Forecasting.

Shown in Figure 6, the parameter estimates analysis has estimated Intercept = -10.43 (downward trend slope). Though, the t test P-value = 0.18 > 0.05 (could not reject the Null Hypothesis of Intercept= 0). There are two possible reasons of not rejecting the Null Hypothesis: (1) there is a strong seasonal component existing in the Glacier Mass data. In the Non-Seasonal ARIMA model, this strong Seasonal component signal would be treated as Non-Seasonal Noise and weaken the Signal-Noise Ratio in Parameter Estimate t test, and (2) the sample size may not be sufficient. Glacier data was collected in 2002-2021 (20 years). If the seasonal component is very strong (12 months), then 20 years of sample size (signal) may not be sufficient as compared to 12 months Seasonal (Noise) in the Non-Seasonal ARIMA model. Authors would continue addressing this subject in Section 4 Seasonal ARIMA model.

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate	Mu
Intercept	0	-10.42635	7.758467	-1.34	0.1807	-10.426349	-10.426349

Figure 6. Non-Seasonal ARIMA Parameter Estimates.

3.4 Non-Seasonal ARIMA Forecasting

In Section 3.3, the forecasting would follow a straight line of slope = -10.43 GT/Month which would indicate that our earth will lose another 2,500 GT (Giga Ton) in next 20 years if the melting not getting faster as shown in Figure 7. Current Both Non-Seasonal ARIMA (1,1,1) and I (1) forecasting models Hyndman (2008, 2018) could not consider any non-linear long-term decaying mechanism. Therefore, there is not much benefit to discuss the AR and MA components for non-seasonal ARIMA model. The Forecasting may be good for long-term trend behavior but missing the short-term Seasonal component month-month pattern within each year. The non-Seasonal ARIMA model could not carry the Seasonal component in the Forecasting model. The Seasonal ARIMA model would be addressed next in Section 4.

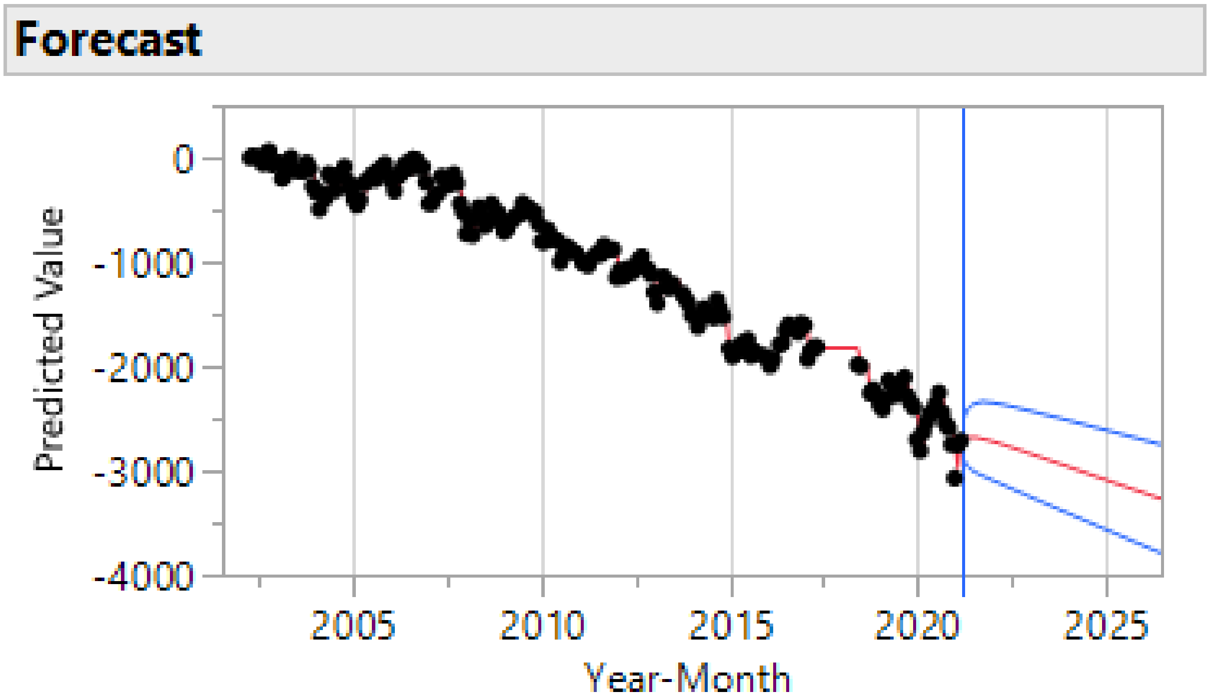

Figure 7. Non-Seasonal ARIMA Forecasting

4 Time Series Seasonal ARIMA Model

Two major concerns were discussed in previous Section 3 Non-Seasonal ARIMA forecasting model: (1) potential long-term trending, (2) missing the month-month Seasonal component. In Section 4, Seasonal ARIMA model would further address these two concerns.

4.1 Seasonal ARIMA Algorithm

In addition to Non-Seasonal ARIMA model, the Seasonal ARIMA model has added the Seasonal Component as (p, d, q) (P, D, Q) m. (P, D, Q) is based on the Seasonal pattern. m= 12 here is representing 12 months in a season (year). For example, in previous ARIMA (1,1,1) model, "d=1" means the trend component d=1 is a straight line in Forecasting. "d=1" is the differencing (delta) is constant between any two consecutive months, resulting in a constant slope of linear trend. In Seasonal ARIMA model, the "D=1" component would compare the same month of two consecutive years (season = one year). This "D=1" component in the Seasonal ARIMA model could detect any year-to-year non-linear long-term trend in addition to non-Seasonal "d=1" linear trend model.

4.2 Seasonal ARIMA Model Group List

To simply the Seasonal ARIMA model list for model comparison, as shown in Figure 8, the Non-Seasonal ARIMA portion has been limited to I (1). Both the AR and MA modules would be addressed in Seasonal portion (P, D, Q) better. Non-Seasonal I (1) was kept in the Seasonal ARIMA model because it may make more sense to consider both the local linear trend of differencing between two consecutive months and the global non-linear trend of differencing between two consecutive years based on 12 months of a Season. Through these model selection list, we may directly compare the strength of the short-term trend Vs. the long term trend through the Seasonal ARIMA Model analysis. The relative strength of these two trend methods may indicate whether the Antarctic Glacier may melt faster in the next 20 years.

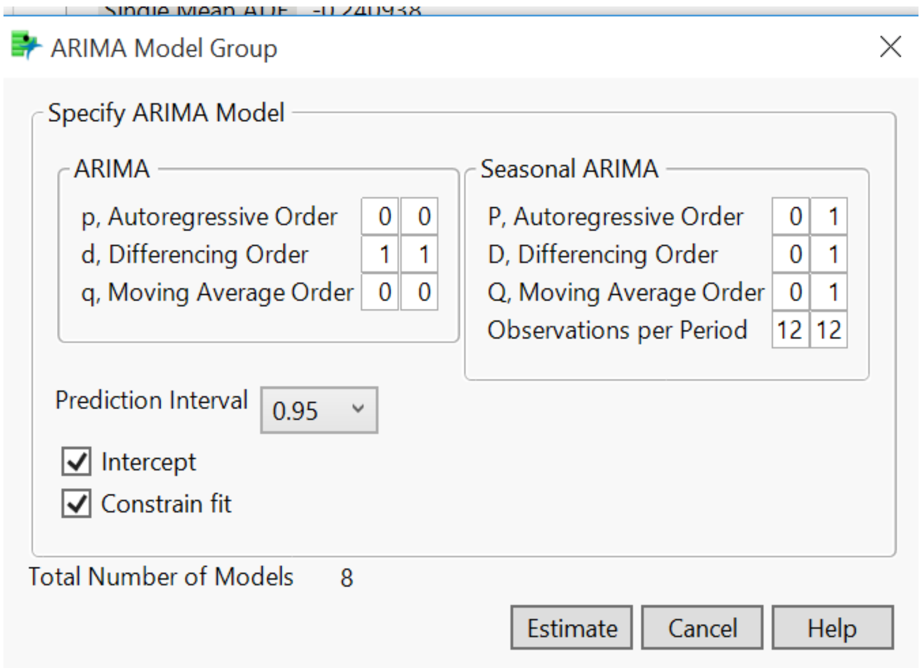

Figure.8 Seasonal ARIMA Model Selection List Menu

Seasonal ARIMA models were ranked based on AIC criteria as shown in Figure 9. ARIMA (0,1,0) (0,1,1)12 was identified as the best model. Previous non-Seasonal I (1) was on the bottom. This new Seasonal ARIMA model may indicate four major findings: (1) Seasonal component is very strong in Antarctic Glacier Melting forecasting, (2) Non-Linear long-term trend "D=1" is significant, (3) Autoregression can be ignored in ARIMA, and (4) Moving Average method is necessary in ARIMA. Authors would address these major findings later in Section 4.

Model Comparison

Report	Graph	Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	Weights
☑	☑	Seasonal ARIMA(0, 1, 0)(0, 1, 1)12	142	10764.574	1752.4447	1758.3843	0.987	1748.4447	0.719820
☐	☐	Seasonal ARIMA(0, 1, 0)(1, 1, 1)12	141	10849.195	1754.4228	1763.3322	0.987	1748.4228	0.267725
☐	☐	Seasonal ARIMA(0, 1, 0)(1, 1, 0)12	142	11610.319	1760.5585	1766.4981	0.986	1756.5585	0.012455
☐	☐	Seasonal ARIMA(0, 1, 0)(0, 1, 0)12	143	13807.891	1782.4022	1785.3720	0.984	1780.4022	0.000000
☐	☐	Seasonal ARIMA(0, 1, 0)(1, 0, 1)12	175	8310.8597	2124.4716	2134.0170	0.988	2118.4716	0.000000
☐	☐	Seasonal ARIMA(0, 1, 0)(1, 0, 0)12	176	9575.9072	2140.2722	2146.6358	0.987	2136.2722	0.000000
☐	☐	Seasonal ARIMA(0, 1, 0)(0, 0, 1)12	176	9992.2064	2147.1390	2153.5025	0.987	2143.139	0.000000
☑	☑	I(1)	177	10785.478	2159.0394	2162.2212	0.986	2157.0394	0.000000

Figure 9. Seasonal ARIMA Model Comparison

4.3 Analyze Seasonal I and MA Modules

In Figure 10 Parameter Estimates, the MA module t test was significant (P-value < 0.05) and the Intercept Trend component was relatively weaker (P-Value > 0.05). The significant MA term may indicate the importance of the smoothing out the random error noise for forecasting in the Seasonal ARIMA model Shiskin (1967). Even the nonlinear long-term intercept is not significant, the yearly decaying slope is still -2.38 GT/year as compared to -10.43 GT/month or -125 GT/year. Even with less than 2% contribution of this non-linear trend term, after 10 years, the contribution or impact of the Forecasting accuracy will be near 20% (faster Glacier melting rate than the non-seasonal forecasting). Therefore, the Seasonal ARIMA model has significantly upgraded the forecasting power of the long-term Glacier Forecasting.

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
MA2,12	2	12	0.632670	0.100669	6.28	<.0001*
Intercept	1	0	-2.381520	4.317669	-0.55	0.5821

Constant Estimate	Mu
-2.3815202	-2.3815202

Figure 10. Seasonal ARIMA Parameter Estimate.

Even with less than 2% contribution of this non-linear trend term, after 10 years, the contribution or impact of the Forecasting accuracy will be near 20% (faster Glacier melting rate than the non-seasonal forecasting). Therefore, the Seasonal ARIMA model has significantly upgraded the forecasting power of the long-term Glacier Forecasting. As shown in the Figure 11, both non-seasonal and seasonal ARIMA forecasting were side by side compared. The Seasonal ARIMA model has shown the seasonal pattern and faster decaying trend than the non-Seasonal model.

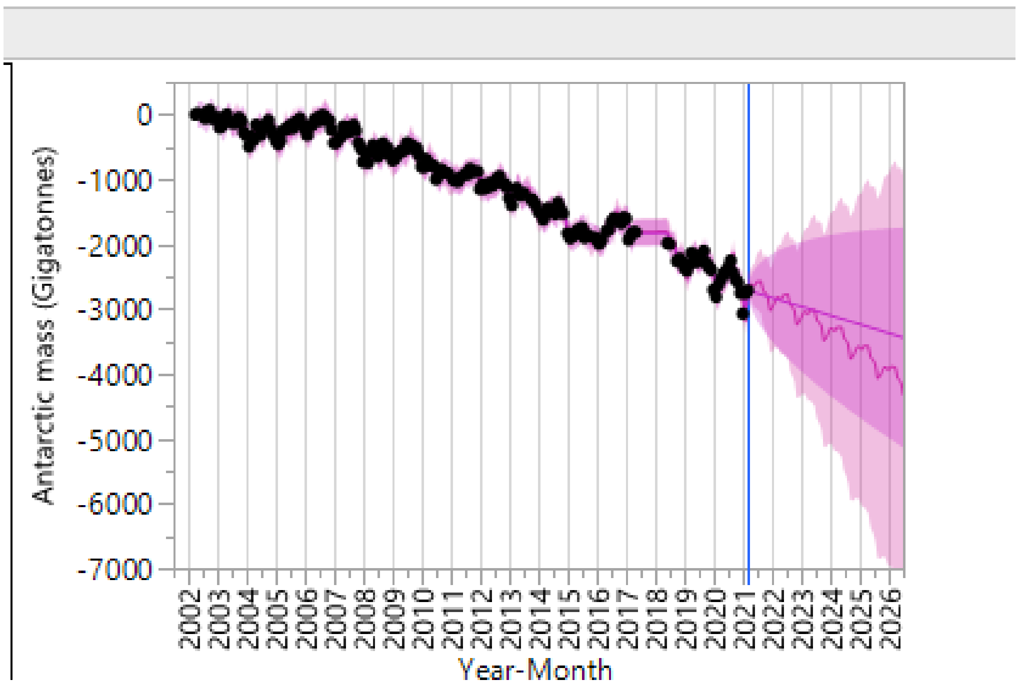

Figure 11. Compare Forecasting between Non-Seasonal and Seasonal ARIMA Models

4.4 Seasonal ARIMA Forecasting

To further investigating the Seasonal ARIMA Forecasting model for next 20 years (2021-2041), JMP Time Series Forecasting platform was conducted and the result was shown in Figure 12. The Seasonal Model Type was identified as AMM (Additive for Error Type, Additive for Trend Type, Multiplicative for Season Type). The details of the JMP ARIMA Model Type won't be addressed here. Readers may refer to the JMP Document Library PDF. The Model report has plotted the Forecasting trend and the 95% prediction interval range. Based on the Seasonal ARIMA forecasting, the Antarctic Glacier would lose another 3,000 GT in next 2021-2041 which is higher than previous Non-Seasonal ARIMA at 2,500. This 20% faster Glacier Melting rate is consistent than the previous section 4.3 parameter t test and intercept. Though, there are two other factors which may impact the Glacier melting rate in 2021-2041: (1) the COVID-19 factor: data included analysis in this paper is from 2002 April to 2021 March including the on-going COVID-19 period. COVID-19 may have significantly limited human activity and reduced air pollution, and (2) the current Seasonal ARIMA model was limited to order of one. MA module could smooth out the error term exponentially by considering more weight on the most recent data such as 2020-2021. The long term yearly differencing trend may consider the faster melting rate as the first order. Authors are considering both the COVID factor and "D=2 or 3" to address any potentially faster glacier melting rate in 2021-2041.

Model Reports

Antarctic mass (Gigatonnes)

Model Type

Model Abbreviation	AAM
Error Type	Additive
Trend Type	Additive
Damped	No
Season Type	Multiplicative
Season	12
Constrained	Yes

Model Report

Series and Forecasts

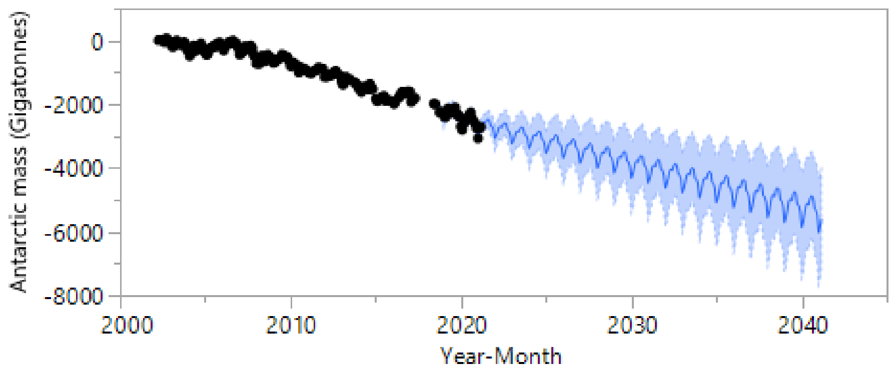

Figure 12. Seasonal ARIMA Model Type and Forecasting for 2021-2041.

5 Conclusions
Exponential Smoothing techniques could decompose the time series components and enhance the Forecasting Power Non-Seasonal ARIMA model is working best when no seasonal component existing and it can forecast the linear or non-linear trend pattern. Seasonal ARIMA is working much effectively when a strong seasonal component existing and it can carry both seasonal component, local linear trend, global non-linear trend in Forecasting future points. Searching the optimal Seasonal ARIMA model could be done by today's JMP platform by setting the selection criteria. Seasonal ARIMA model can forecast at about 20% faster Glacier Melting Rate than the non-Seasonal ARIMA model for 2021-2041.

Future Work
Authors are continuing current Antarctic Glacier project: study the COVID-19 factor, learn more Adv. Time Series Techniques such as State Space Smoothing, Higher Order Seasonal ARIMA Models, Forecasting and Prediction Interval Statistics…

Acknowledgements
Thanks to our JMP Advisor Patrick Giuliano and IEOM STEM Co-Chairs Dr. Ali and Dr. Reimer.

References
Arnold, Denis G. (ed.), *The Ethics of Global Climate Change*, New York: Cambridge University Press, 2011.
Baillie, RT, Chung, SK, "Modeling and forecasting from trend-stationary long memory models with applications to climatology", *International Journal of Forecasting*. Volume 18, Issue 2, Pages 215-226, 2002.
Bindoff, Nathaniel L., Peter A. Stott, et al.,"Detection and Attribution of Climate Change: from Global to Regional", in Stocker et al. 2013: 867–952, 2013.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.
Burnham, K. P., and Anderson, D. R. "Multimodel Inference: Understanding AIC and BIC in Model Selection." *Sociological Methods and Research* 33:261–304, 2004.
Burnham, K. P., Andersen, D. R., and Huyvaert, K. P. "AIC Model Selection and Multimodel Inference in Behavioral Ecology: Some Background, Observations, and Comparisons." *Behavioral Ecology and Sociobiology* 65:23–35, 2011.
Hyndman, R.J., Athanasopoulos, G. *Forecasting: Principles and Practice*, 2018.
Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D. *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin: Springer-Verlag. 2008.
Multivariate SPC Methods for Process and Product Monitoring." *Journal of Quality Technology* 28:409–428.
Shiskin, J., Young, A. H., and Musgrave, J. C. *The X-11 Variant of the Census Method II Seasonal Adjustment Program. Technical Report 15*, US Department of Commerce, Bureau of the Census, 1967.

Biography

Mason Chen is a junior high school student in the Online High School of Stanford University (SOHS). Chen has certified IASSC Black Belt, IBM SPSS Statistics, Modeler Data Mining, and JMP STIPS certificates. Chen has also published more than 50 papers in the STEAMS (Science, Technology, Engineering, AI, Mathematics, Statistics) and has won many awards in IEOM STEM and Six Sigma Competitions. Chen has found his STEAMS Organization (website: stem2steams.weebly.com and also found the SOHS STEAMS Club since his middle school years. SOHS club has sent many students to IEOM and complete well in most STEM competition events. Mason Chen is currently concentrating on the Computational Biology research fields.