TABLE OF CONTENTS

CLAIM-COUNT-FORECASTING

ABC insurance firm specializes in multiple products across the personal and commercial insurance spectrum. This company's market share ranking is within the top 10, regardless of product offering, industry, or market. The goal of this project is to analyze and forecast ABC's claim frequency for their domestic (USA) commercial property portfolio.

Commercial property is a very difficult line to understand. It is generally characterized as a severity line, meaning infrequent claims throughout a given year but when claims happen, they can be worth $1M-$30M per claim. Property claims also have extreme volatility due to catastrophic events like wildfires, hurricanes, tornados, etc.… as well as less extreme weather events. Other than weather, Fire remains the largest single peril for ABC and the industry. Fire losses are typically by business practice like industry type or behavioral characteristics like management, employee training, or equipment/facility practices. Due to this volatility, predicting claim amounts is very difficult and requires different modeling techniques like capital tranching or excess load modeling. Understanding claim volumes is an important step within the actuarial process to generate a prospective opinion on reserve amounts for the line.

BUSINESS CASE

Review, analyze, and build a forecasting model(s) to predict a year's worth of claim volume. This forecast will be used by the reserving actuaries in their quarterly reserve review process. Look for logical clusters or groupings to improve model predictions. Understand if there are trends in claims over time as well as seasonal patterns in the perils or industries. Last, determining how catastrophic events influence the claim forecasting methods.

DATA

Data set from ABC Insurance. The dataset represents 10+ years of commercial property claims. These claims represent all possible loss types, like fire, wind, water, hail, hurricanes, etc.… and multiple industries. ABC Insurance has products for small commercial exposures like a coffee shop or a single bay mechanic shop all the way through to complex large commercial exposure like multi-state/internal operations or large single buildings. All the data has been masked to protect the confidentiality nature of the information, including loss dollars. In the raw output the loss dollars have a formula applied to change the values but retain the patterns and relationships to each other. In the accompanying zip file, please review the files "Claim_Modeling_Data_Extract" and "Claim_Modeling_Data_Extract_Final".

Data Attributes:

"Claim_Modeling_Data_Extract.xlsx"

Column Names	Description	Data Type
U_ID	Unique key used to identify a claim and a coverage type. Has been masked from original data	Numerical (int); primary key
LD_ID	Date of claim made	Date (YYYYMMDD)
C_ID	Unique number assigned to claim. Has been masked from original data	Categorical
S_ID	State where the claim occurred. Has been masked from original data	Categorical
Z_ID	Zip Code where the claim occurred. Has been masked from original data	Categorical
C_ID	Catastrophe code assigned by PCS. has been masked from original data	Categorical
CAT_ID	Catastrophe indicator Y/N	Categorical
P_ID	Policy number for claim submitted. Has been masked from original data	Categorical
R_ID	Reporting group (aka business unit). Has been masked from original data	Categorical
PG_ID	Program code which is used to group similar type risks like manufacturing or retail. Has been masked from original data	Categorical
E_ID	Policy effective date	Date (YYYYMMDD)
SI_ID	Code used to associate exposure to an industry. This differs from PG_ID because industry codes are external and PG_ID is internal to the insurance company	Categorical
F_ID	Policy form code which is used to determine which product had the loss. Has been masked from original data	Categorical
CV_ID	Coverage code used to determine what was damaged and applicable terms and conditions. Has been masked from original data	Categorical
WW_ID	Weather Water identification to group water losses	Categorical

Column Names	Description	Data Type
LG_ID	Loss group description used to describe the peril that caused the loss, example Fire. Has been masked from original	Categorical
LGD_ID	Loss sub group description used to describe the peril that caused the loss in greater detail, example lightening. Has been masked from original	Categorical
W_ID	If the peril that caused the loss was Weather then it will be Y else it will be N	Categorical
BLD_ID	If the claim and the coverage has building damage then it will be Y else it will be N or it could be mixed	Categorical
Loss_ID	Total loss amount (paid + incurred). Has been modified from the original data by applying arithmetic to the original values	Numeric

Within the excel “Claim_Modeling_Data_Extract" file there is a tab called “Counts” (hidden) where all of the following numbers come from. In total there are 223,984 claims for the entire dataset, of which 4,799 claims are from the developing year of 2021. These records were removed from the data set because 2021 is a developing year and there are additional data challenges with developing claims. Example, a claim can occur on 1/1/2021 but it could take the insured 20-30 days to contact the insurance company because the insured did not know there was significant damage like in the case of hail or ice damming. Using data as new as 2020 as of mid-2021 is a more developed data set.

After removing 2021 claims, the dataset shrank to 219,185 claims (233,984-4,799). As a point of reference 219,185 claims divided by 10 years is about 21,919 then cutting the data by month or 21,919/12 = about 1,827 per month or even smaller at about 59 (1,827/31) per day. This is a very small volume of claims, so the team aggregated the data primarily using domain knowledge. The following 5 key aggregation assumptions are made:

(1) Aggregate claims on a monthly basis: Given our business challenge is to assist the reserving actuaries, daily claim forecasting is not as important because reserves are set monthly and quarterly. With quarterly reserves seen by Wall Street analysts and monthly used as an internal temperature check. Claim patterns tend to be based on weather activity so there is a strong seasonal component, which monthly will represent nicely.

(2) Aggregate all of the geographic variables into regions: Geography plays an important role with claim activity and since the data is masked a proper mapping table will not be provided, instead the mapping was done when the excel “Claim_Modeling_Data_Extract_Final” file is created. The regions selected represent the company’s major sales regions and these regions already take into consideration how the business is transacted and the type of risks. 5 regions were created: West, South, South Central, North East, North West, and Upper Midwest.

(3) Aggregate all of the market lines of business and products into 4 groups: Small, Middle, Large, and Other: Since the market sizing speaks to the types of products offered, the risks appetite, and the

underlying exposure complexity, this was a logical group to make. Similar to point 2 about a proper mapping table will not be provided, instead the mapping was done when the excel "Claim_Modeling_Data_Extract_Final" file is created. Small markets typically insure 1 - 2 locations, have total insured values less than 10 million, and range from agriculture, service, and office type risks. Middle markets typically have 10+ locations, have total insured values 25M+, and range from retail and wholesale trade, manufacturing, services, and office risks. Large markets typically have 50+ locations with total insured values of 50M+ and range from habitational risks, universities/colleges, high-rise buildings, malls/stripmalls and service/retail industries. "Other" is our excess and surplus lines which can be any risk within the company's underwriting appetite that the admitted market is unwilling to write. These risks can range from small, middle, or large markets but have higher potential risk and are thus unregulated by the various departments of insurance due to the coverage restrictions and need for flexible pricing. Example, coastal or earthquake prone areas, high risk industries like fertilizer manufacturers, or capital impaired companies.

(4) Aggregate industries (great reference: https://www.osha.gov/data/sic-manual) into 7 divisions based on domain knowledge from ABC Insurance: Similar to point 2 about a proper mapping table will not be provided, instead the mapping was done when the excel "Claim_Modeling_Data_Extract_Final" file is created. The 7 divisions are Agriculture, Forestry, And Fishing, Finance, Insurance, And Real Estate, Manufacturing, Other, Public Administration, Retail and Wholesale Trade Services.

(5) Aggregate claims into catastrophes (CATs) vs non-CATs: CATs are a major driver of claims for any property insurance company. CATs are unpredictable due to their infrequent nature or their ability to impact many insureds in a single event. A single CAT also has a wide range of severity uncertainty, they can drive a handful of very expensive losses or a lot of insignificant losses or both. CATs are defined by Property Claim Services (PCS), which is an ISO company and is used as an industry standard. PCS's definition is any event causing $25M+ of insured property damage and impacting multiple insurers. These events are typically hurricanes/tropical storms, earthquakes, terrorism, wildfires, hail, straight-lined winds (derechos), winter storms, flooding, etc.… Given the role that CATs play on insurance claims, Y/N indicator using the C_ID field has been created.

In addition to the data modifications mentioned above, retained only the following fields: W_ID, BLD_ID, & LG_ID and threw away the rest. The new attributes for modeling exercise can be found in the excel "Claim_Modeling_Data_Extract_Final" file. Below is a new table used to start our modeling exercise:

Data Attributes: "Claim_Modeling_Data_Extract_Final.xlsx" tab "Con_Project_Data"

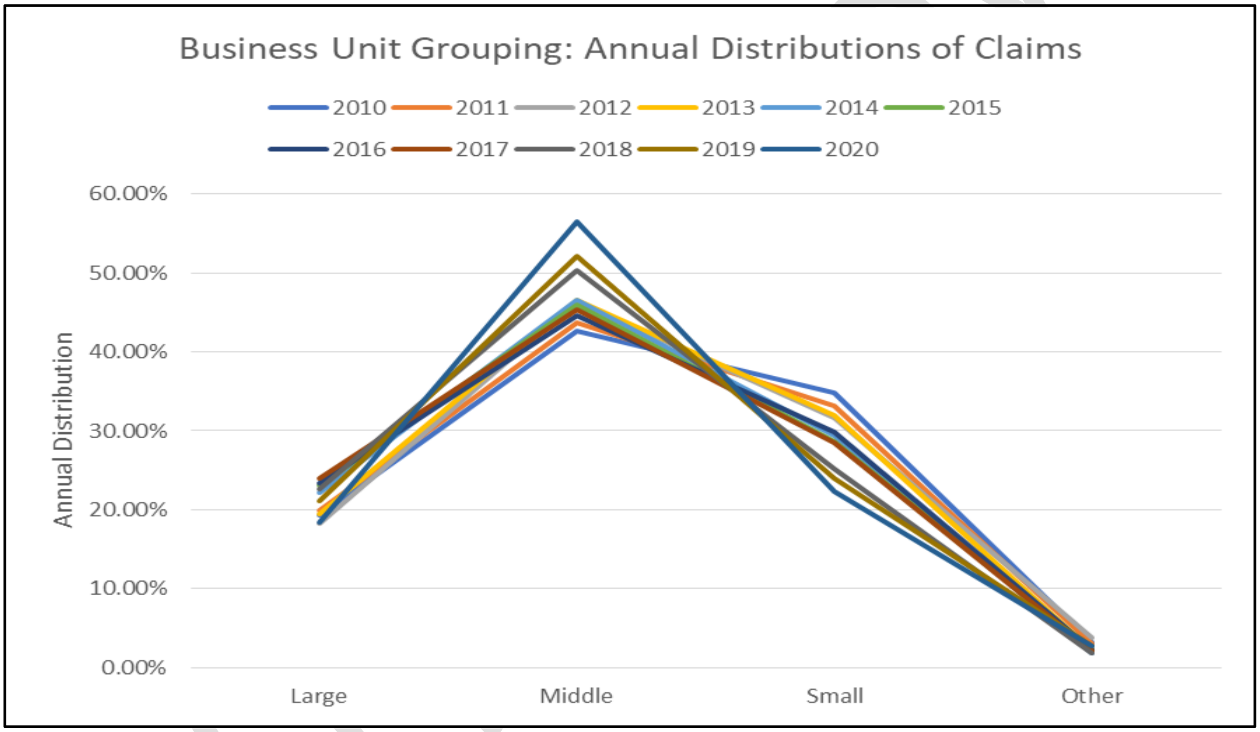
Column Names	Description	Data Type
Date_of_Claim_Mnthly	Translated daily claims from LD_ID into month and year. Defaulting the day to the first day of the month	Date (YYYYMMDD)
Region	Translated state to region, explained above	Categorical
CAT_IND	Translated C_ID to Y=CAT claim and N=Not a CAT claim, explained above	Y/N
Market_Grouping	Translated R_ID and F_ID into market groups: Small, Middle, Large, Other. Explained above	Categorical

Column Names	Description	Data Type
Industry	Translated PG_ID and SI_ID into 7 industry groups: Agriculture, Forestry, And Fishing, Finance, Insurance, And Real Estate, Manufacturing, Other, Public Administration, Retail and Wholesale Trade Services. Explained above	Categorical
Water_From_Weather_ID	Renamed WW_ID from original extract. No other change. Y=Water claim was from weather and N=Water claim was not from weather	Y/N
BLD_ID	Taken from the original extract no change. If the claim and the coverage has building damage then it will be Y else it will be N or it could be mixed	Categorical
Loss_Group	Taken from the original extract, no change. Loss group description used to describe the peril that caused the loss, example Fire. Has been masked from original	Categorical
Weather_ID	Taken from the original extract, no change. Y=Claim was from a weather event, N=Claim was not from a weather event	Y/N
Total	Accumulation of transactional claim data to a count of claims	Numerical (Target variable)
Year	Not used in modeling, rather information field to do exploratory work in excel and it is derived from the attribute Date_of_Claim_Mnthly	Numerical
Peril	Not used in modeling, rather information field to do exploratory work in excel and it is derived from the Loss_Group attribute	Categorical

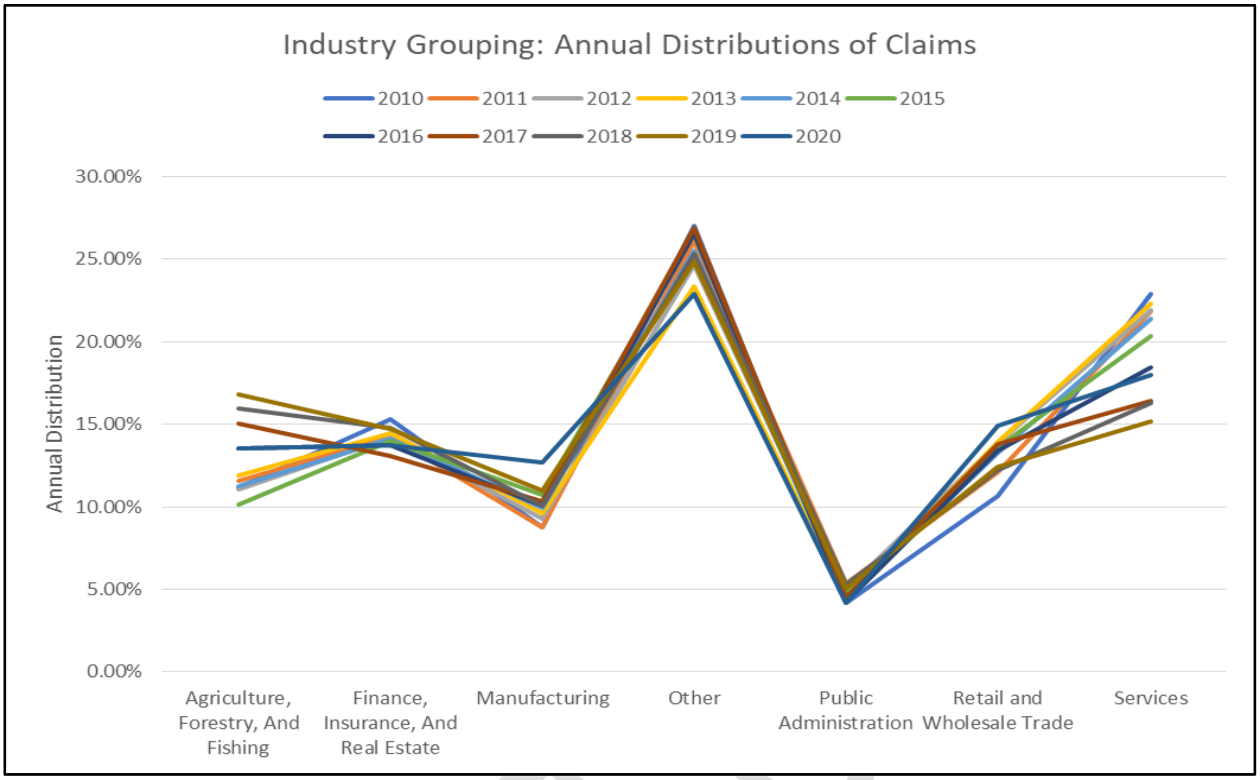
DATA EXPLORATION:

Before loading the data into SAS, the data has been explored via excel to understand what types of models to build and if there was anything "interesting" in the data. All the proceeding images and data points can be found within the "Claim_Modeling_Data_Extract_Final.xlsx".

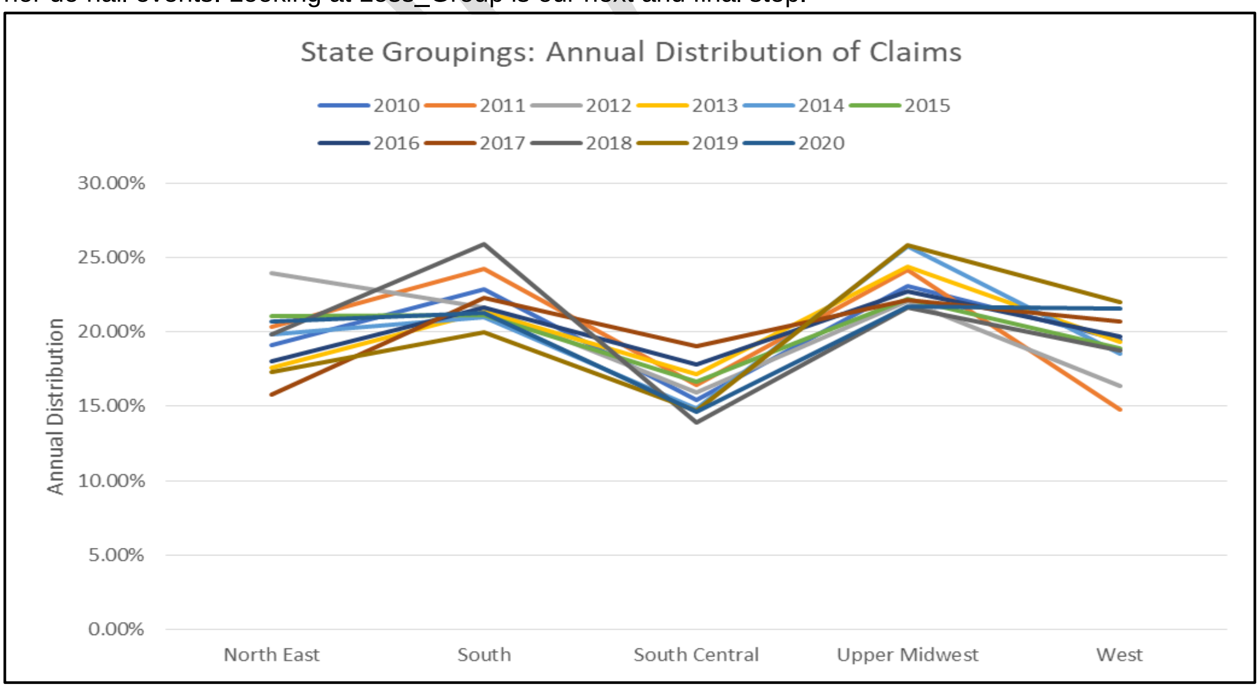
The team analyzed 4 different distribution graphs to understand if the different attributes were going to be helpful at predicting future claim volumes. Our goal was to see if different attributes influenced the annual claim distribution thus resulting in a need to go further. Starting with Marking groups, year to year the range by group is tight. Large and Other generally make up 20% and 3%, respectively of the annual claim volume. Middle and Small fluctuate a little more with 47% and 29%, respectively of the annual claim volume with a few years higher or lower but relatively in-line. There could be something to this attribute or it could be another signal, which will be explained below.



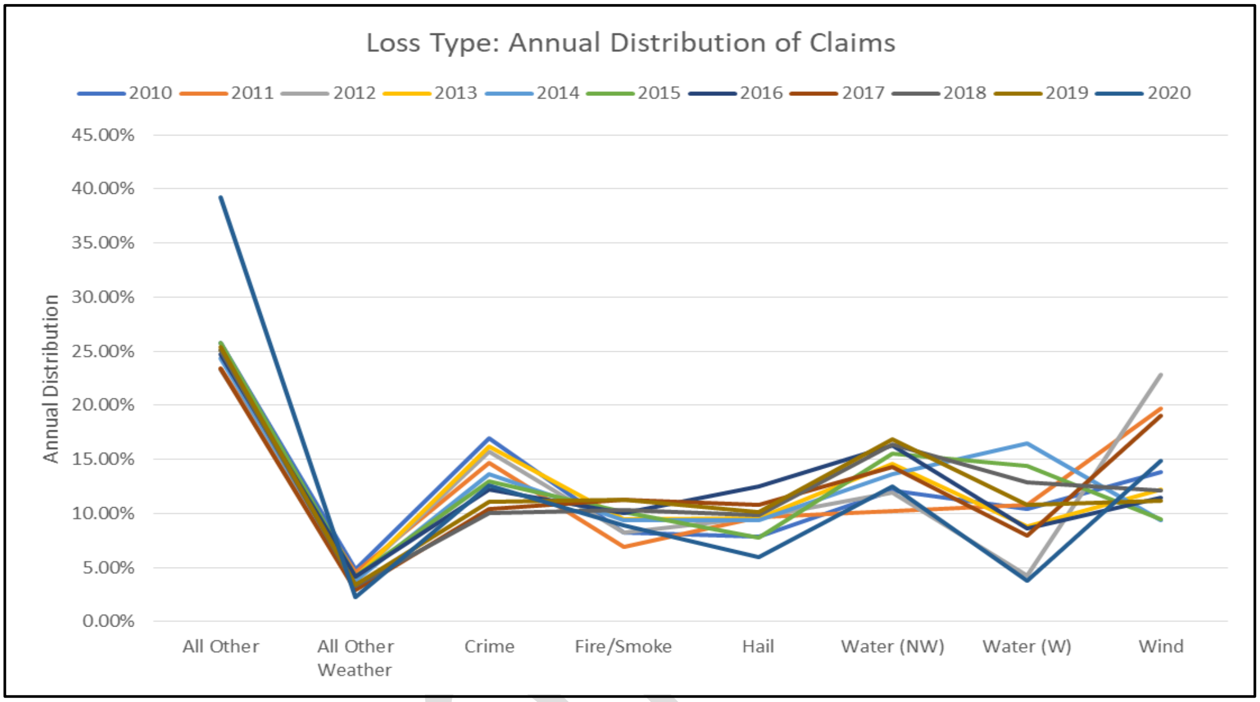
Next we looked into industry groups. Here we see another similar graph as above. There are a few industries that are changing their mix each year like agriculture or services but the majority of these industries are stable. Manufacturing tends to represent 10% of the claims, finance/insurance/real estate tends to represent about 14% of the claims, and other tends to represent 25% of the claims. Thus, there could be something here but likely not.



Next we look at the 5 regions and this is where it starts to be more apparent on what is going on within the data. Each year the mix of regions keeps changing and since regions are great explanations for weather and CAT activity. The only issue with regions is it is too broad and a single region can have events hit multiple exposures within the region and outside the region. Think hurricanes do not stop at state lines nor do hail events. Looking at Loss_Group is our next and final step.



Property claims are heavily influenced by the weather. In the last 10 years weather claims have represented anywhere from 37% - 45% of ABC’s claim volume. As can be seen on the below chart. Only 2 perils are constant, “All Other” and “All Other Weather”. The other perils vary year over year and make up a different percentage of the overall claim activity.



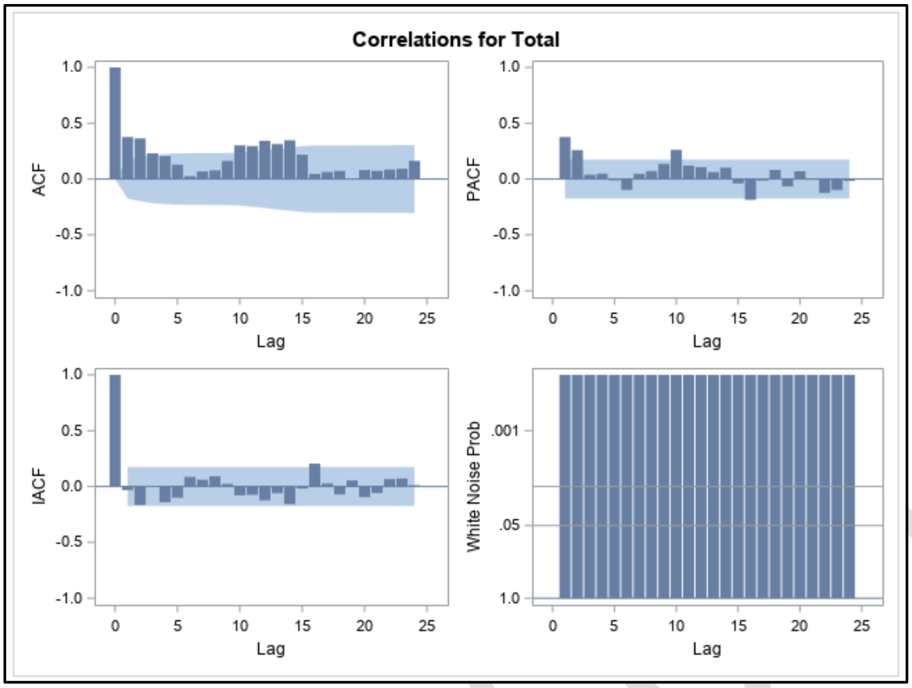
What can be modeled?

Now shifting to SAS, we continued to explore the data set “Claim_Modeling_Data_Extract_Final.xlsx” to design our modeling approach. To keep the data simple and easy to work with the team extracted the tab “Con_Project_Data” into its own workbook called “SAS_Import.xlxs”

Importing to SAS: Using the SAS import feature, the following code was generated. Once the table was imported it was moved into the library and renamed to “project”.

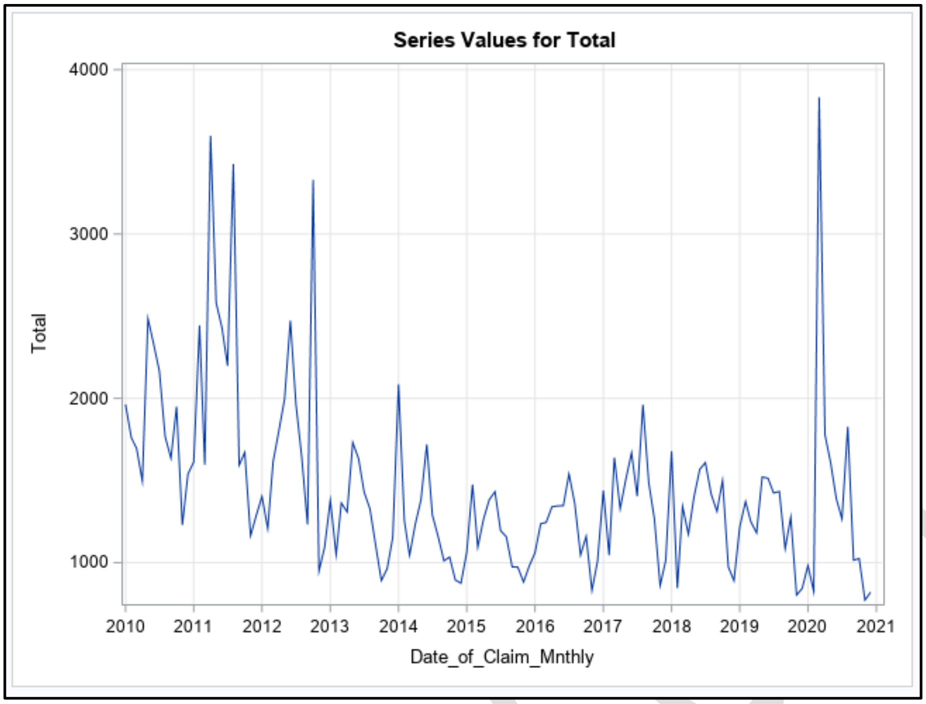
Once in SAS the team explored the data to check for autocorrelation and white noise. To do this we used the forecasting task called “Time Series Exploration”. The task was set up to use “Total” (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and the transformation of “Total” to accumulation by means of summation. We ran a gambit of plots and reports.

In total we can see that there is autocorrelation as shown on the ACF and PACF plots and there is a signal given the white noise plot the tall bars tell us to reject the null hypothesis for the white noise test.



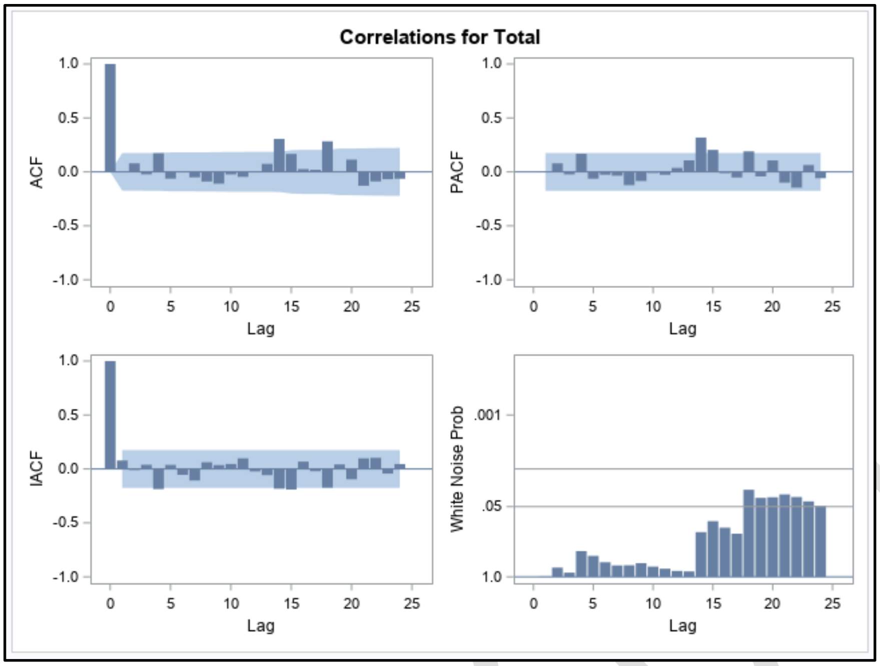
The chart below also shows that there could be irregularities or events that will need attention. Given the spikes occurred in 2011, 2012, & 2020, this finding led the team to explore if CAT activity is the driver. Sure enough, winter storm Alfred and the worst tornado/hail season took place in 2011, Sandy happened in 2012, and COVID in 2020 (reference: https://www.natlawreview.com/article/pandemic-vs-policyholder-covid-19-and-business-interruption-coverage-claims).

Given that CATs are hard to predict and prior CAT activity has no bearing on a future CAT activity, it is reasonable to believe that CAT activity might be white noise and lack the proper signal for time series modeling. With the one caveat, Hail activity is more predictable because it happens every year and there are many events that take place in a given year, which is very different from other CAT activity like hurricanes or wildfires.



To look at CATs excluding hail we used the forecasting task called "Time Series Exploration". Within this task we used the following filters: CAT_IND = 'Y' and Loss_Group <>7. The task was set up to use "Total" (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and the transformation of "Total" to accumulation by means of summation. We ran a gambit of plots and reports.

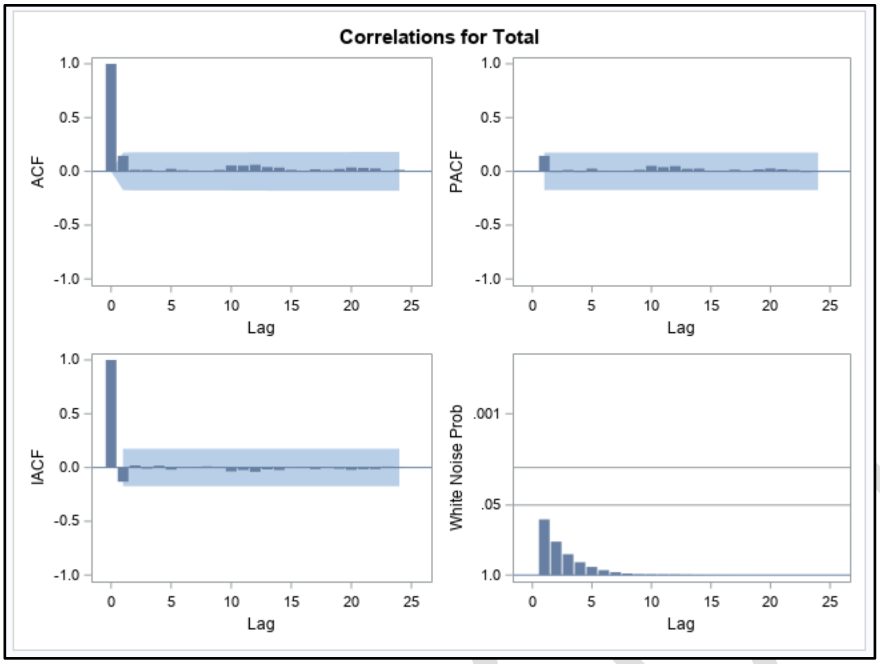
Our assumption holds, the below plot shows that CAT activity excluding hail lacks autocorrelation by the ACF, PACF, and IACF. The White Noise plot also indicates that the data is all noise and lacks strong signals. Our recommendation is to remove CAT excluding hail claims from our data set and look to other modeling techniques.



There is another type of loss within the dataset that is similar to CATs, it is "All Other Loss". This catch all bucket is dominated by motor vehicles striking into properties and causing exterior and interior damage. This type of activity is random and hard to predict or understand. To determine if it should stay within our data we looked at the correlation plots.

To look at "All Other" we used the forecasting task called "Time Series Exploration". Within this task we used the following filters: CAT_IND = 'N' and Loss_Group <>3. The task was set up to use "Total" (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and the transformation of "Total" to accumulation by means of summation. We ran a gambit of plots and reports.

Our assumption holds, the below plot shows that "All Other" activity lacks autocorrelation by the ACF, PACF, and IACF. The White Noise plot also indicates that the data is all noise and lacks strong signals. Our recommendation is to remove "All Other" claims from our data set and look to other modeling techniques or use a simple average to estimate the future. There are no industry models or homegrown models that we are aware of that try to estimate these types of losses.



To summarize, claim data is relatively consistent year over year by industry & market as seen in the above excel graphs. This is due to underwriting practices and the target appetite the company has for each industry & market. Regional annual variation looked to be important due to the influence weather has on claims. This led to approach the dataset in 2 ways: (1) a weather model and (2) a non-weather model. Before committing to 2 different independent models, we reviewed the outliers. CATs proved to be too random for time series, so it was removed. However, CAT Hail claims are frequent enough to model so we added another dataset (3) CAT/xCAT Hail. In addition, "All Other" claims were too random for time series, so it was removed. The removed data lacked autocorrelation and is white noise (no signal). With this, the team decided to logically break the data into 3 separate parts.

(1) Hail model, which will exclusively look at Hail both CAT and xCAT because this peril does not care what it is hitting and there is little to no action an insured can do to prevent this. The majority of hail damage is roof and cladding, with geography being very important. The type of structure is irrelevant as damage will occur, for example a small shop or a large building exposed to hail will have losses and file a claim. These differences only matter on the claim dollars, not the propensity to file a claim. In addition, the industry cannot truly tell the difference between cosmetic damage and real damage so if hail marks up a building's exterior an insurance company will pay because defending our contract language in court is very costly and has not gone well for the industry when arguing cosmetic damage.

(2) Weather model (excluding hail), which is predominantly water and wind based claims. Water and wind act differently than hail, there is more that an insured can do to limit the chance to file a claim. Example, proper tree maintenance, HVAC roof equipment installed to code, property snow removal from roof, etc.… This group should have a different cause and effect relationship than the hail group.

(3) xWeather model, which is predominantly fire, crime, and water non-weather (burst pipe). Industry and market annual claim distribution does not fluctuate widely and there is limited claim volume, so not breaking this down by industry or market made sense. These perils are mostly behavior, aging equipment, or operations within the facility. Example is the insured properly

maintaining their facility to prevent fires or water damage. Or an aging HVAC system could fail and cause water damage. Or the insured has poor processes/procedures that can lead to fires (true story I have seen large fire losses because a woodworking company let their employees leave stain rags on the floor in the hot sun near door openings, spontaneous combustion). Our underwriter does a very good job at selecting and managing risk, so this keeps volatility to a minimum.

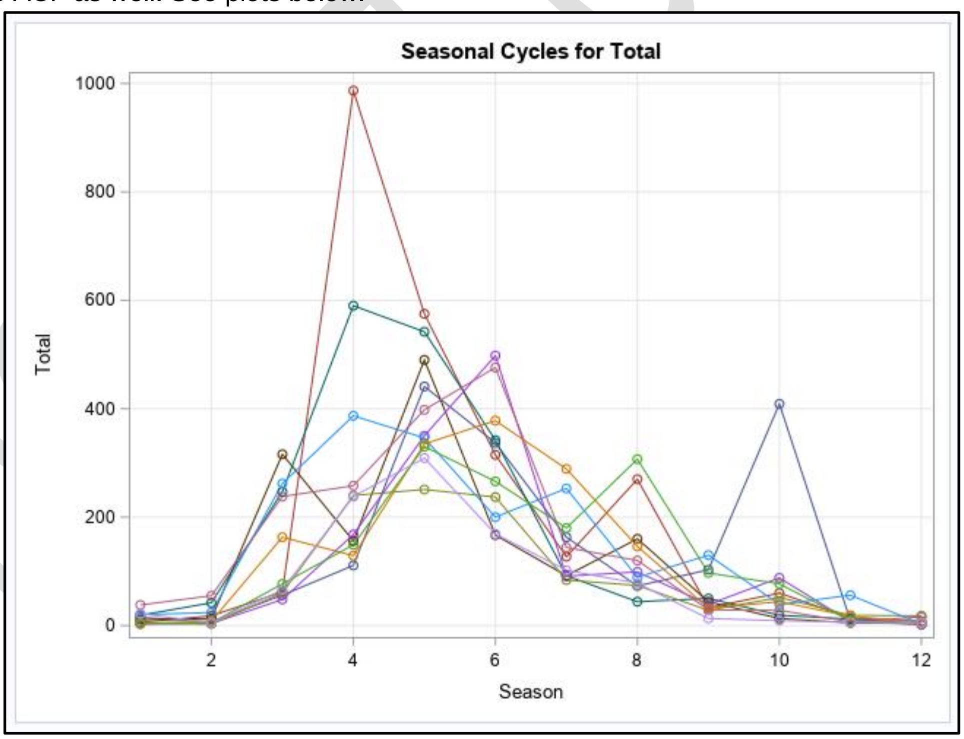
MODELING:

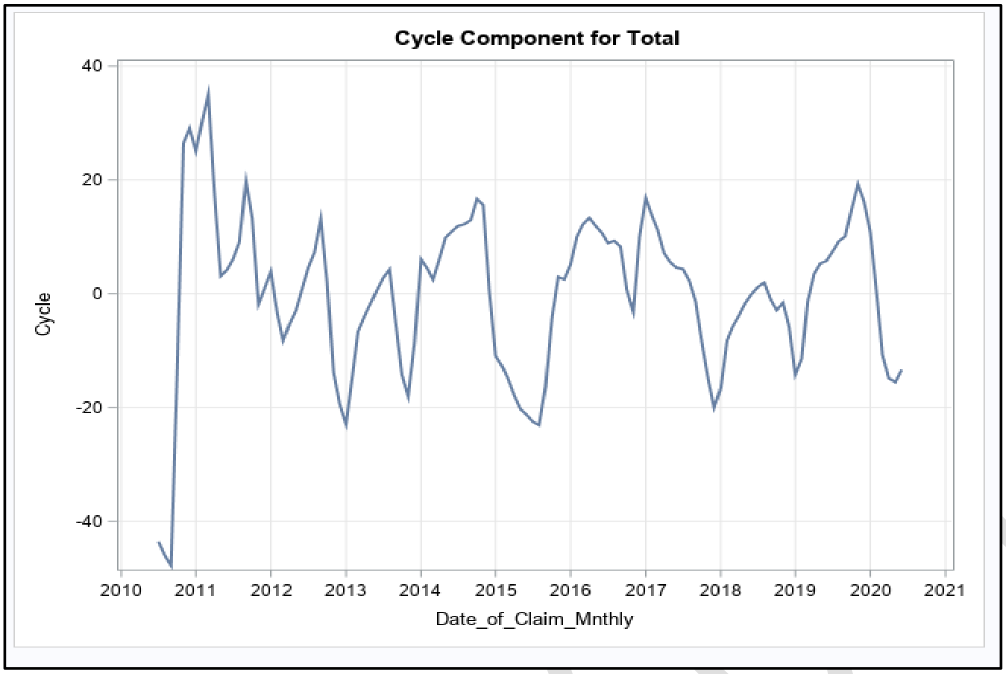
(1) Hail Model:

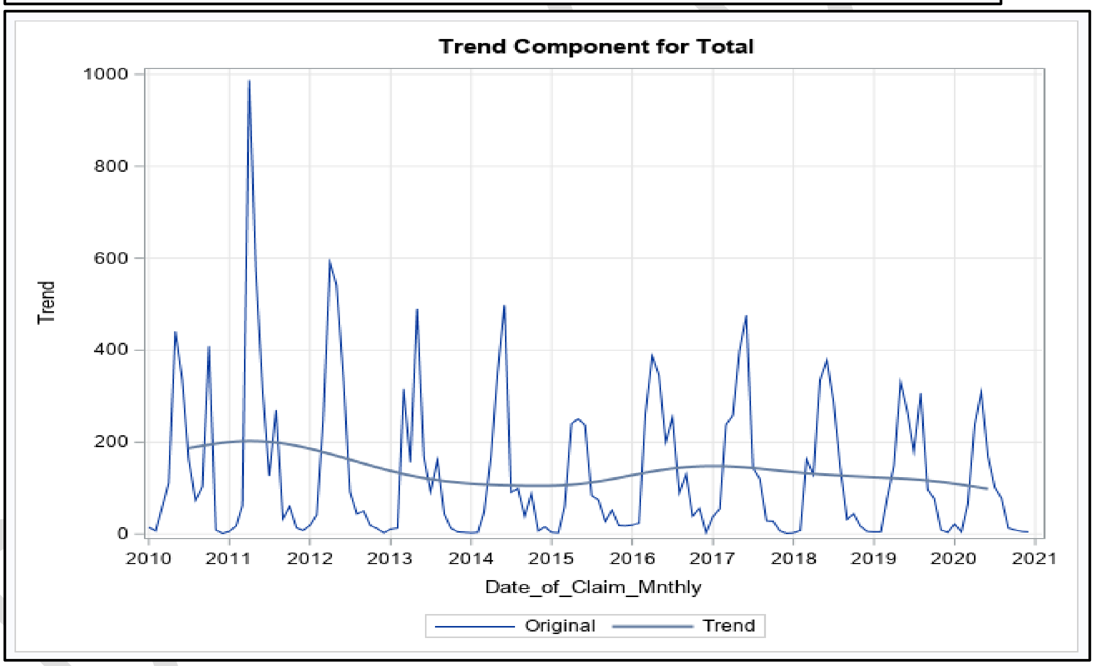
To generate the data needed for a hail model the team used the forecasting task called "Time Series Data Preparation". Within this task we used the following filter: Loss_Group = 7. The task was set up to use "Total" (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and the transformation of "Total" to accumulation by means of summation.

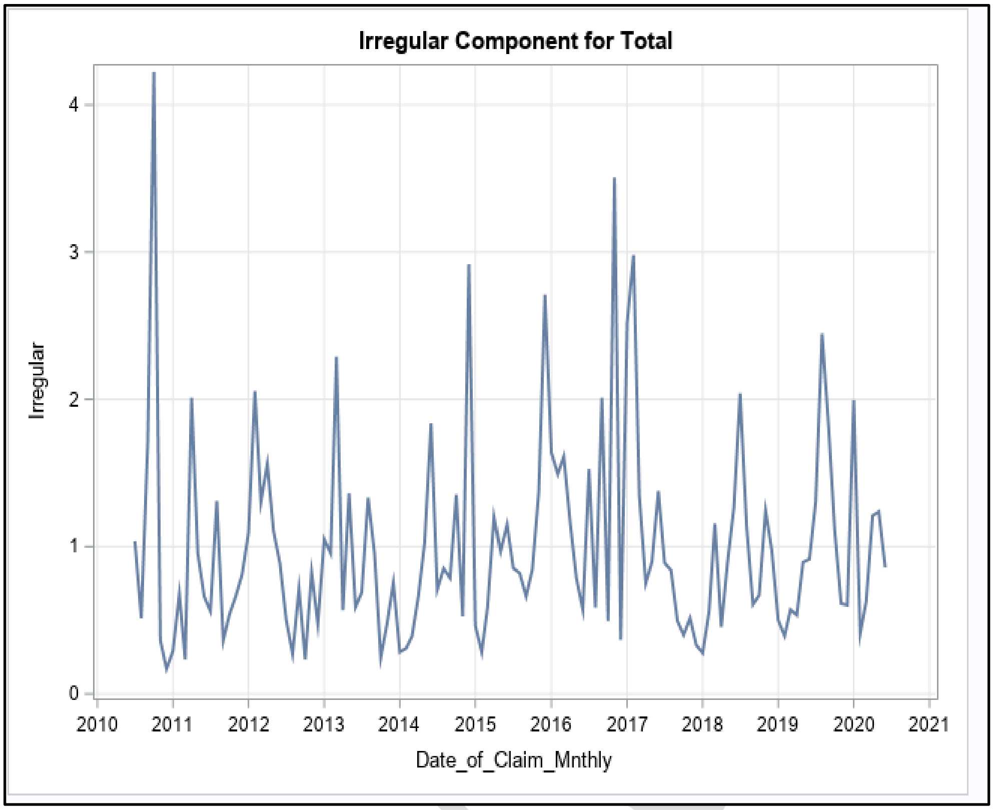
To look at the hail data we used the forecasting task called "Time Series Exploration". The task was set up to use "Total" (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and no transformation as the prior step took care of that. We ran a gambit of plots and reports.

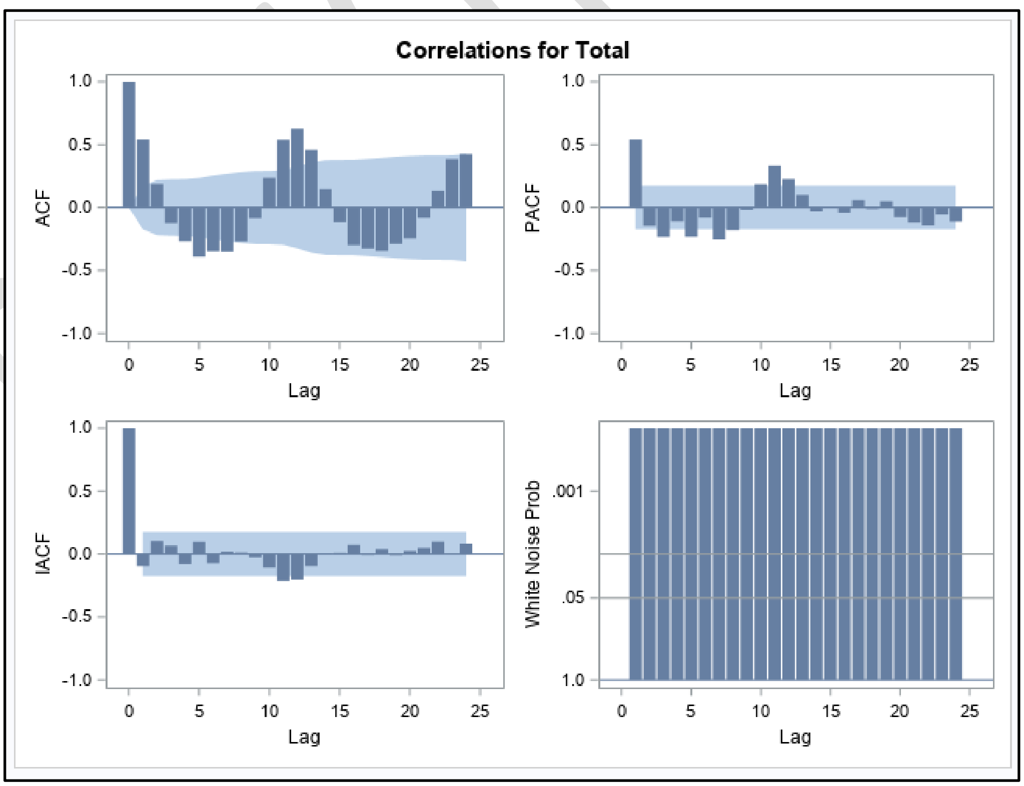
When reviewing this data set 1 thing was apparent and it made sense, hail was seasonal and it happened in the Mar-Jul, see the season plot below. What was interesting is the lack of trend despite the company growing their book of business overtime. This lack of trend could be the direct result of underwriting initiatives aimed to reduce hail exposure due to the costly claims the company has dealt with over the years. These initiatives are increased deductibles, reduced capacity, and co-insurance on buildings with questionable insurance to valuations. All of these initiatives will reduce frequency and thus could explain the lack of trend. Non seasonal cycles do not appear to be an issue nor are their signals form the irregular component. Hail does have autocorrelation and is not white noise. The seasonal patterns can be observed in the ACF as well. See plots below.




Cycle Component for Total
Cycle
40
20
0
-20
-40
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly


Trend Component for Total
Trend
1000
800
600
400
200
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly
Original
Trend


Irregular Component for Total
Irregular
0
1
2
3
4
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly

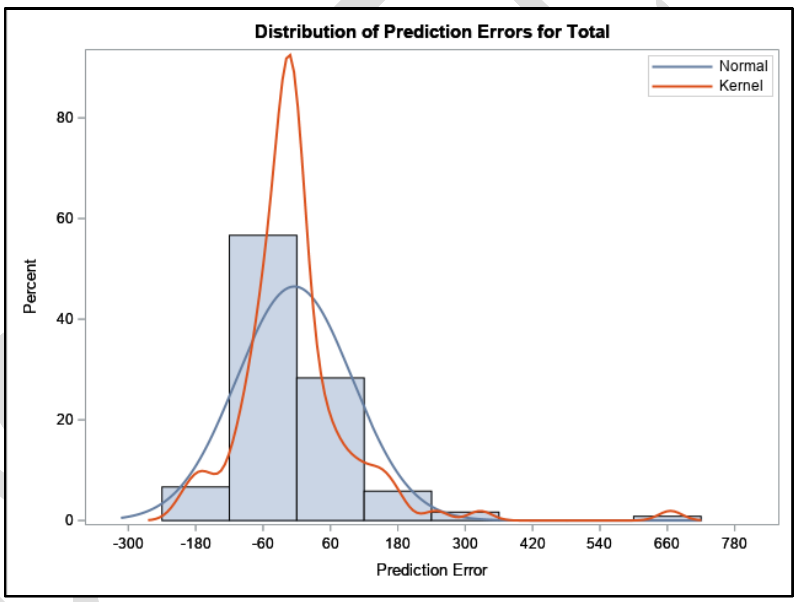

Correlations for Total
ACF
PACF
IACF
White Noise Prob
1.0
0.5
0.0
-0.5
-1.0
.001
.05
0
5
10
15
20
25
Lag

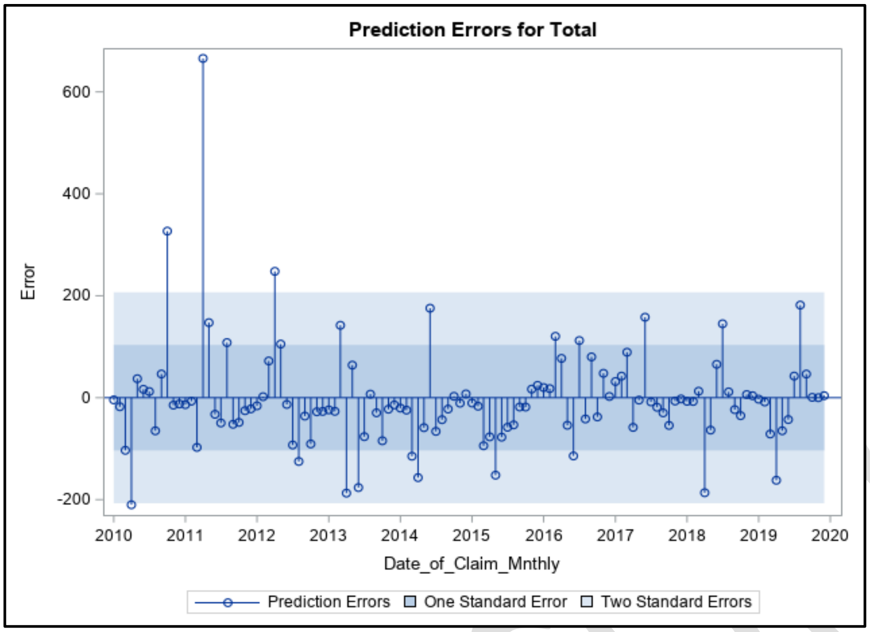
Building the Hail Model:

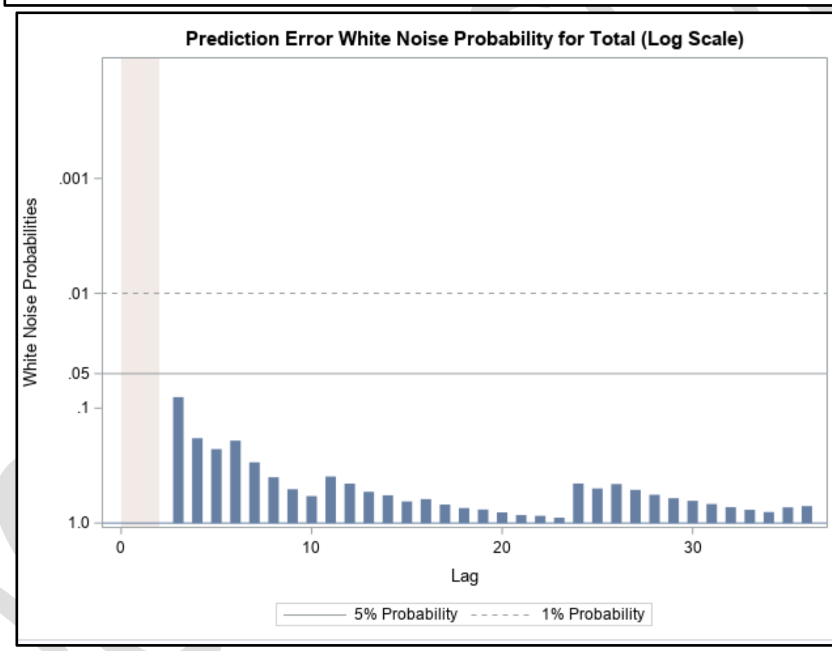
Given there is only seasonality the first 2 models the team explored are (1) Exponential Smoothing Model (ESM) additive seasonal exponential smoothing and (2) ESM multiplicative seasonal exponential smoothing.

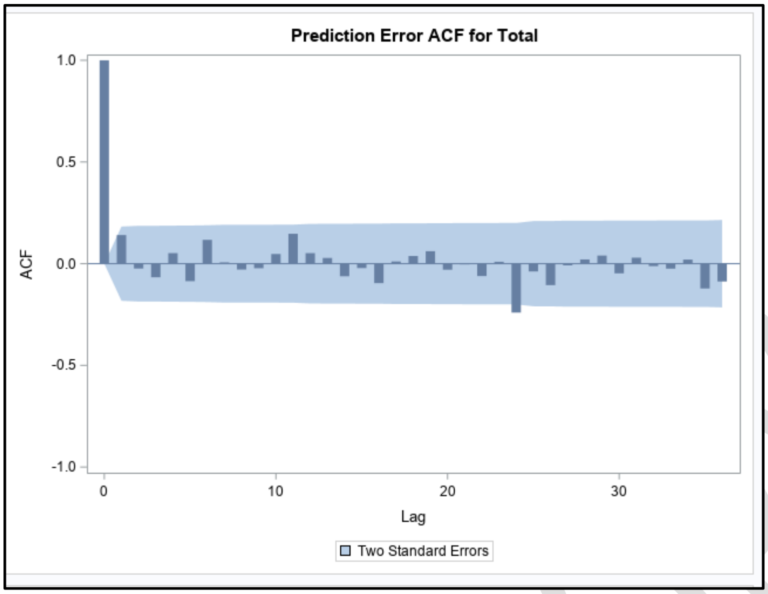
Setting up the ESM additive seasonal exponential smoothing exercise, the team started with the forecasting task "Modeling and Forecasting", selecting the PROJECT_MNTH_HAIL table as the data. "Total" as the dependent variable, date_of_claim_mnthly as the time element. For the model, selected the ESM and selected additive seasonal exponential smoothing and ran all plots. On the options tab, select 12 months to forecast as our data is monthly and 12 months for the hold back. For the output, create fit statics was checked and called it "ESMadd".

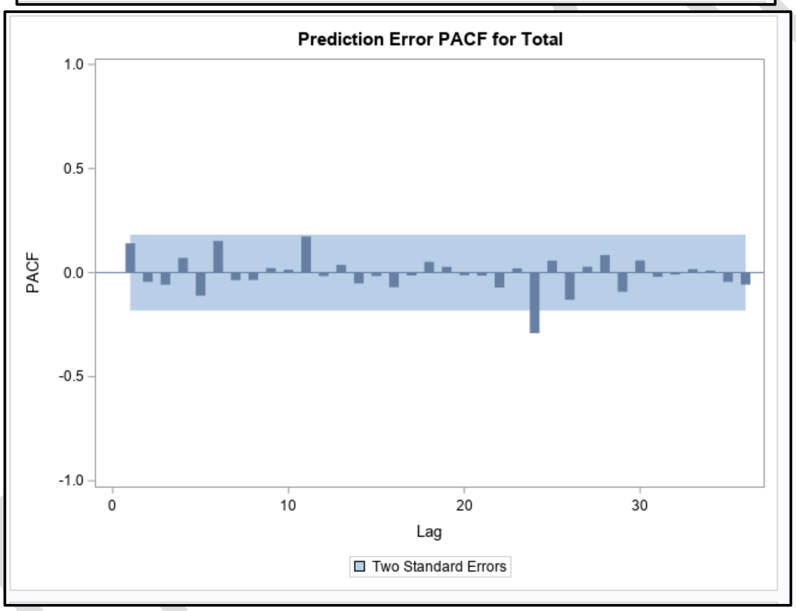
Overall the model did well, errors were normally distributed, mostly within 1 and 2 standard deviations and were White Noise. The ACF and PACF on the errors do not show signs of autocorrelation, with a few lags on the IACF being somewhat significant. When eyeballing the forecast plot the model seems to do a reasonable job at following the seasonality. See plots below. The AIC on the fit was 1,115.45 (rounded) and the SBC was 1,121.024 (rounded). The MAPE and RMSE on the fit were 118.19 (rounded) and 102.62 (rounded), respectively. The AIC on the forecast was 100.59 (rounded) and the SBC was 100.59 (rounded). The MAPE and RMSE on the forecast were 148.13 (rounded) and 66.11 (rounded), respectively.

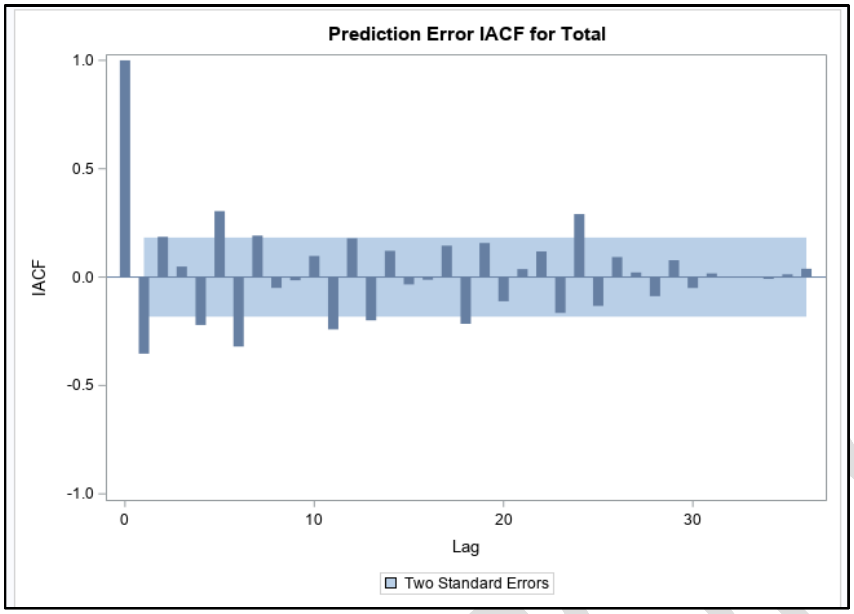


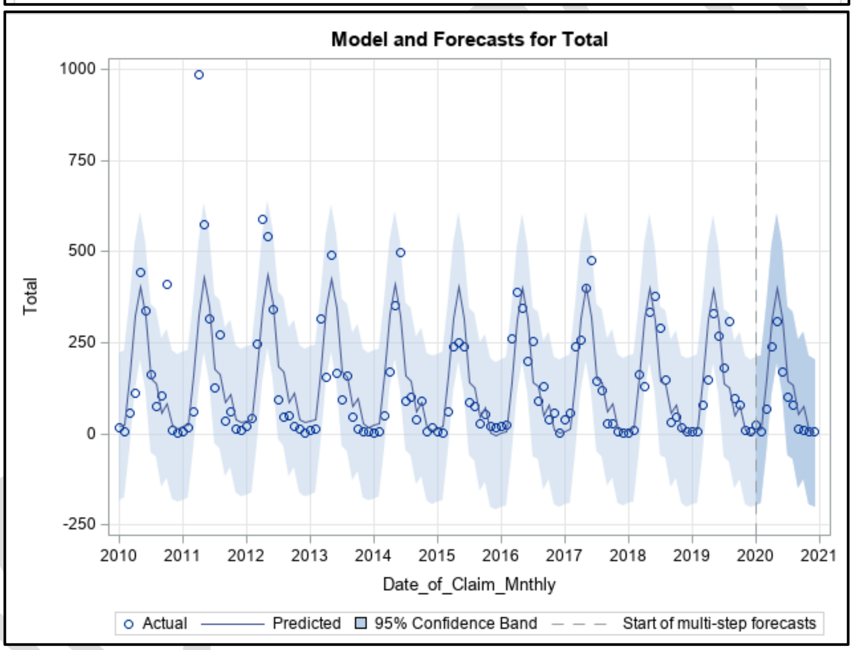

Prediction Errors for Total
Error
600
400
200
0
-200
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Date_of_Claim_Mnthly
Prediction Errors
One Standard Error
Two Standard Errors


Prediction Error White Noise Probability for Total (Log Scale)
White Noise Probabilities
.001
.01
.05
.1
1.0
0
10
20
30
Lag
5% Probability
1% Probability


Prediction Error ACF for Total
ACF
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
Lag
Two Standard Errors


Prediction Error PACF for Total
PACF
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
Lag
Two Standard Errors

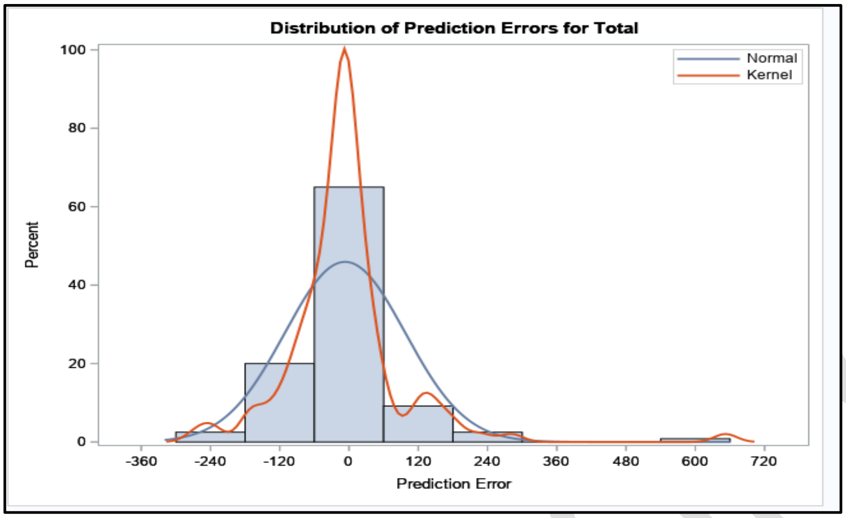


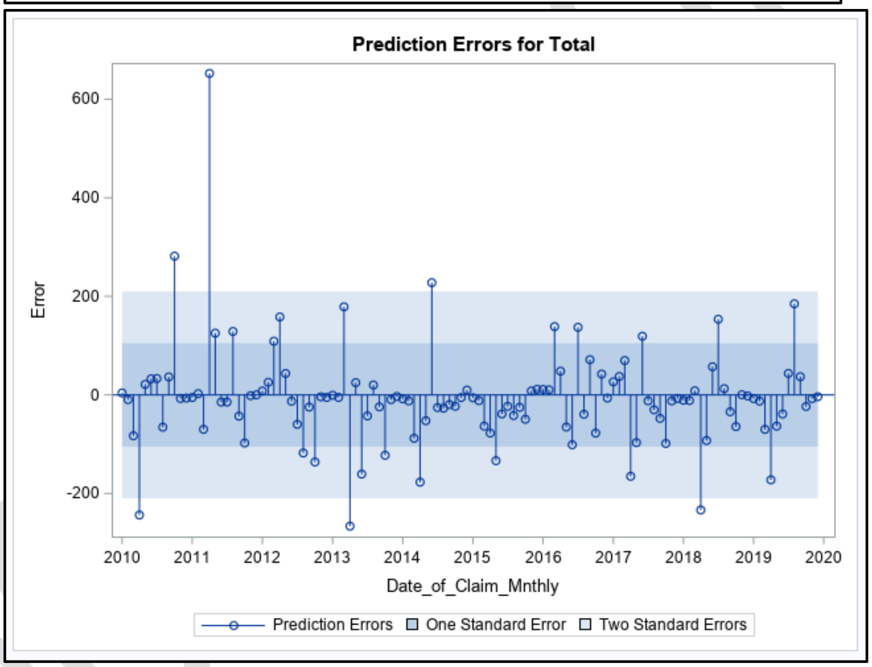


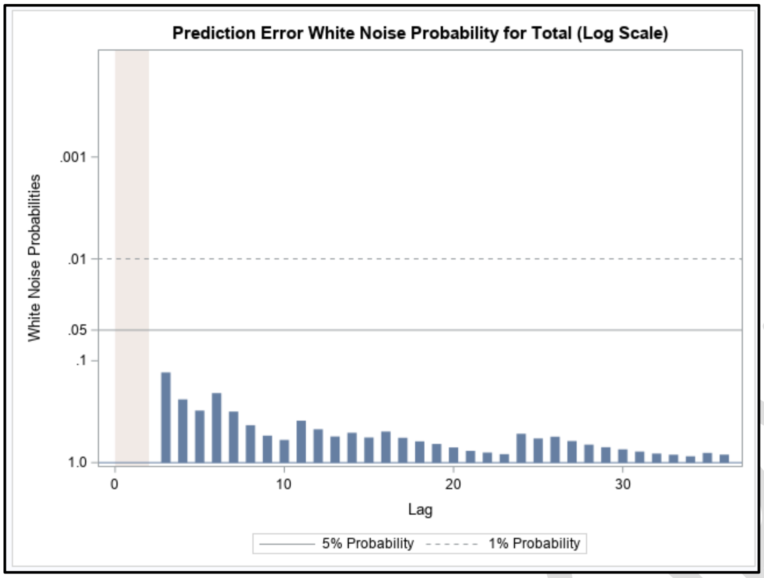
Setting up the ESM multiplicative seasonal exponential smoothing exercise, the team started with the forecasting task "Modeling and Forecasting", selecting the PROJECT_MNTH_HAIL table as the data. "Total" as the dependent variable, date_of_claim_mnthly as the time element. For the model, selected the ESM and selected multiplicative seasonal exponential smoothing and ran all plots. On the options tab, select 12 months to forecast as our data is monthly and 12 months for the hold back. For the output, create fit statics was checked and called it "ESMmult".

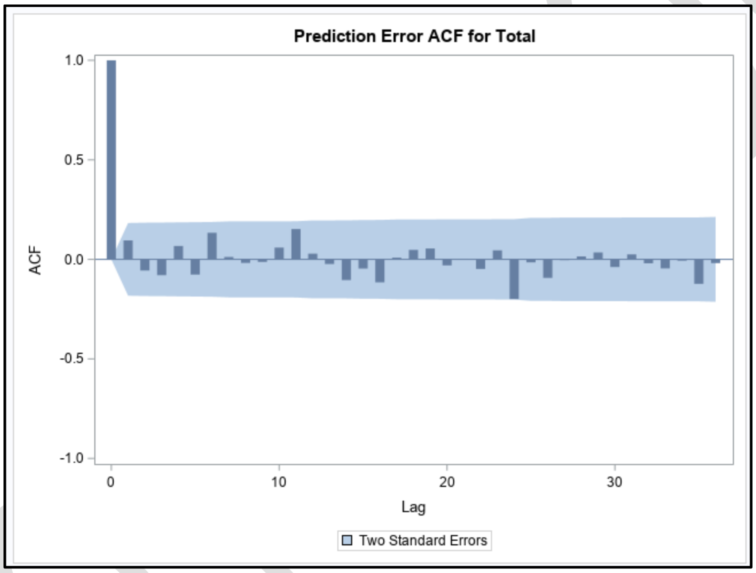
Overall, the model did well, errors were normally distributed, mostly within 1 and 2 standard deviations and were White Noise. The ACF and PACF on the errors do not show signs of autocorrelation, with a few lags on the IACF being somewhat significant. When eyeballing the forecast plot the model seems to do a reasonable job at following the seasonality. See plots below. The AIC on the fit was 1,118.64 (rounded) and the SBC was 1,124.22 (rounded). The MAPE and RMSE on the fit were 96.71 (rounded) and 104.00 (rounded), respectively. The AIC on the forecast was 101.14 (rounded) and the SBC was

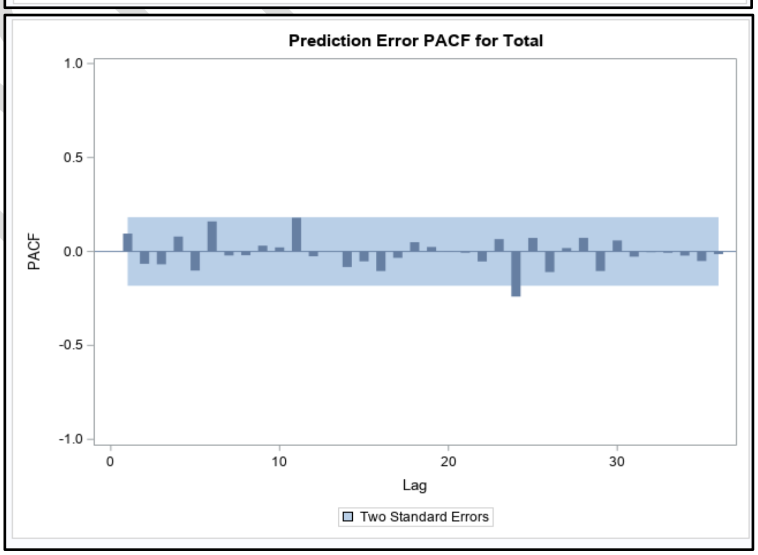
101.14 (rounded). The MAPE and RMSE on the forecast were 182.75 (rounded) and 67.64 (rounded), respectively.

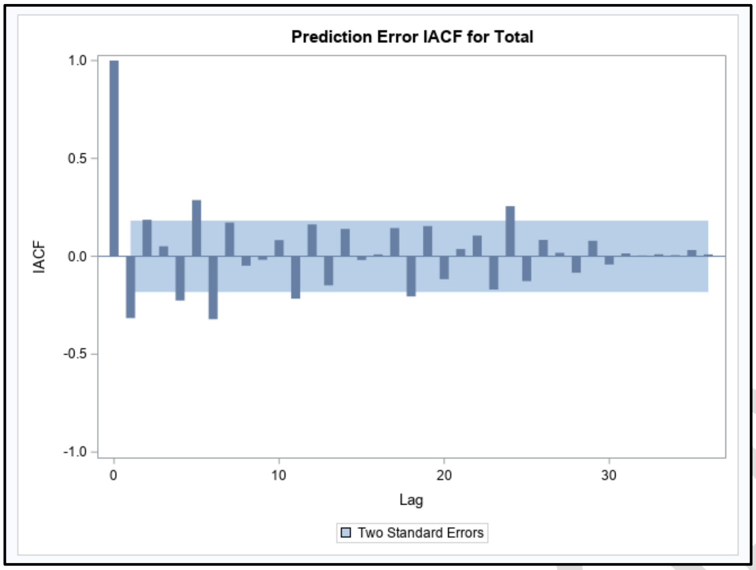


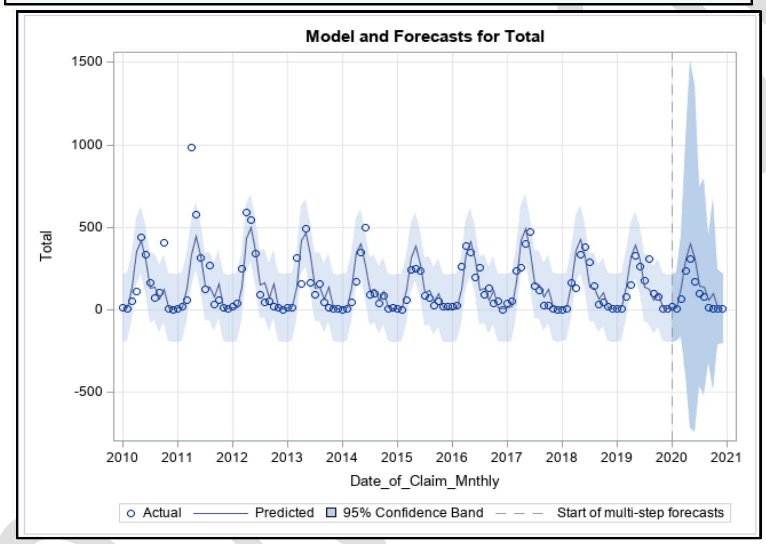



Prediction Error White Noise Probability for Total (Log Scale)
White Noise Probabilities
.001
.01
.05
.1
1.0
0
10
20
30
Lag
5% Probability
1% Probability


Prediction Error ACF for Total
ACF
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
Lag
Two Standard Errors


Prediction Error PACF for Total
PACF
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
Lag
Two Standard Errors

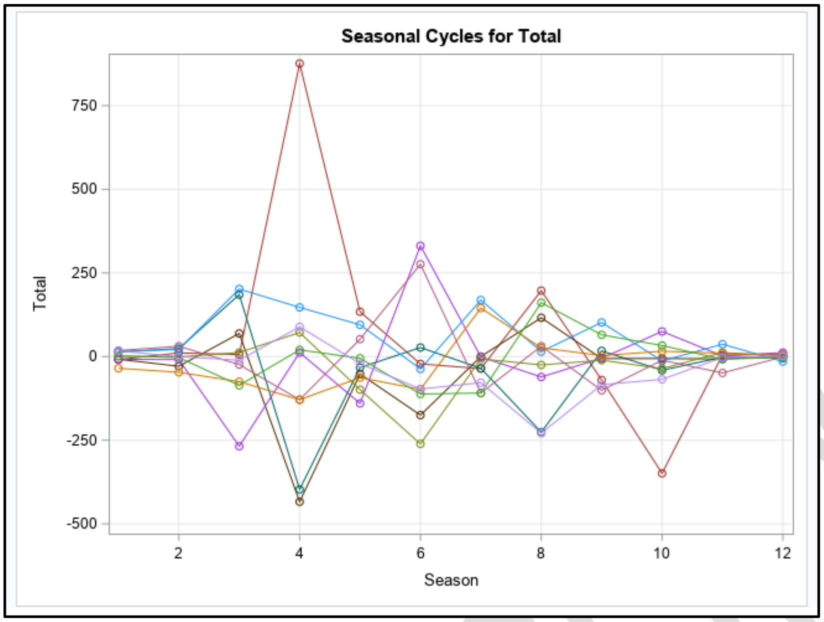


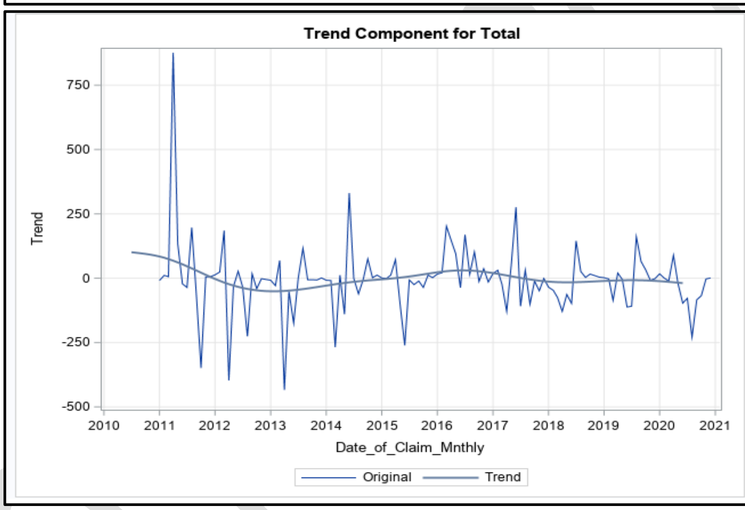


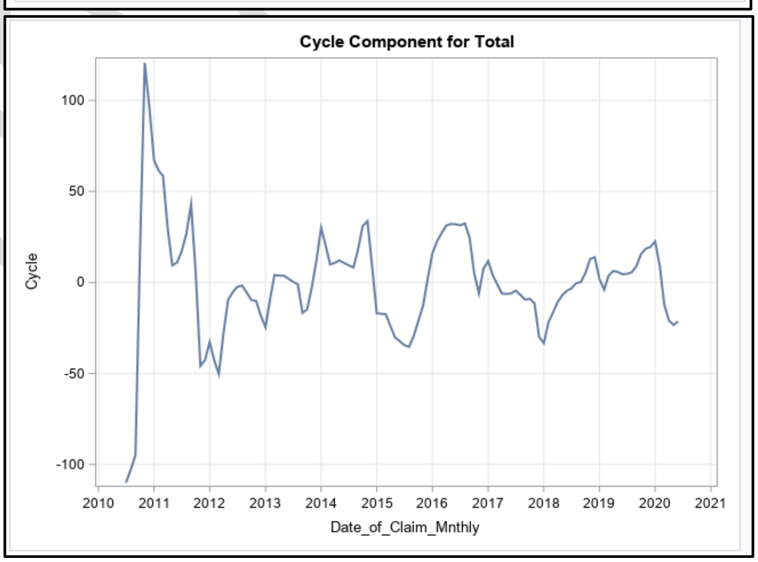
Between the 2 models, **ESM Additive** did a slightly better job and therefore will be the model to be chosen for the Hail model. The ESM additive had a slightly lower SBC & AIC with a slightly worse RMSE and MAPE. However, the ESM multiplicative model had more errors past the 2nd standard deviation. ARMA models were not considered, after transforming the data via 1 seasonal differencing, the data turned into White Noise, thus preventing the use of an ARMA model.

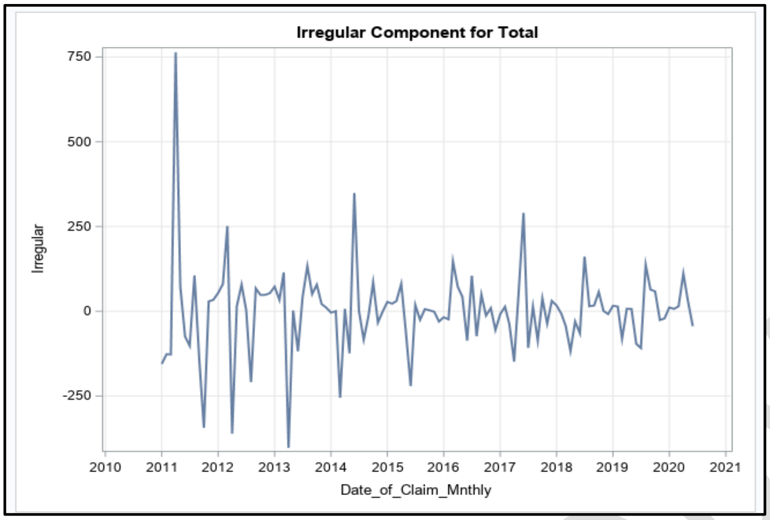
To look at the result of 1 seasonal differencing on the hail data we used the forecasting task called “Time Series Exploration ". The task was set up: “Total” (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and 1 seasonal difference. We ran a gambit of plots and reports.

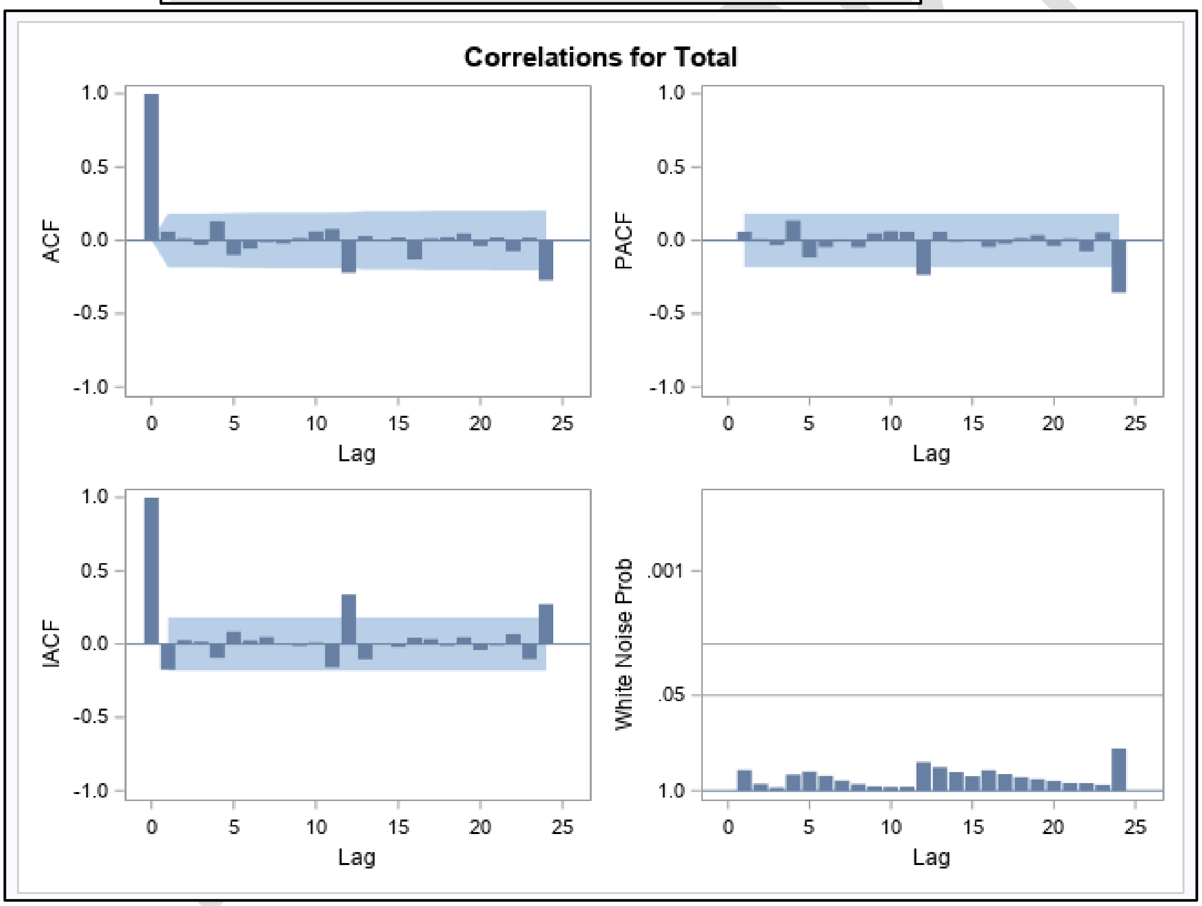
It can be seen that the seasonal differencing took care of the seasonality, there is no trend, cycle or irregular component. The correlation plots do not show autocorrelation and the data is now white noise. See plots below.


Seasonal Cycles for Total
Total
750
500
250
0
-250
-500
2
4
6
8
10
12
Season


Trend Component for Total
Trend
750
500
250
0
-250
-500
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly
Original
Trend


Cycle Component for Total
Cycle
100
50
0
-50
-100
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly


Irregular Component for Total
Irregular
750
500
250
0
-250
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly

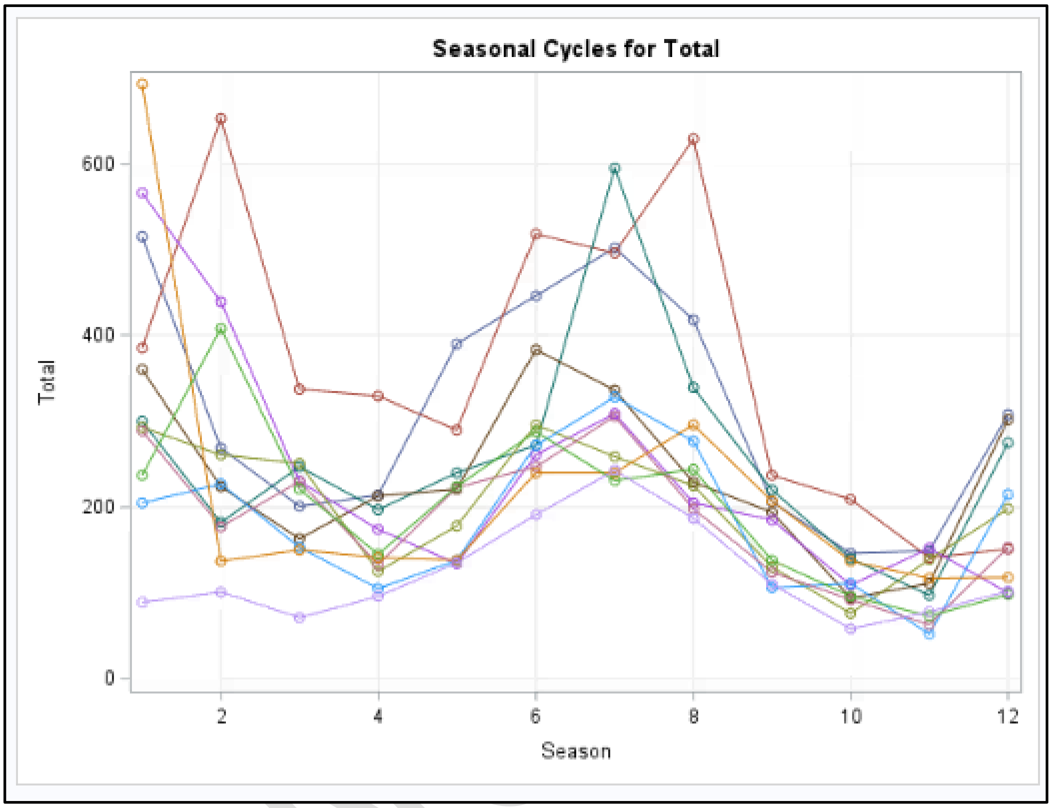

Correlations for Total
ACF
PACF
IACF
White Noise Prob
1.0
0.5
0.0
-0.5
-1.0
.001
.05
0
5
10
15
20
25
Lag

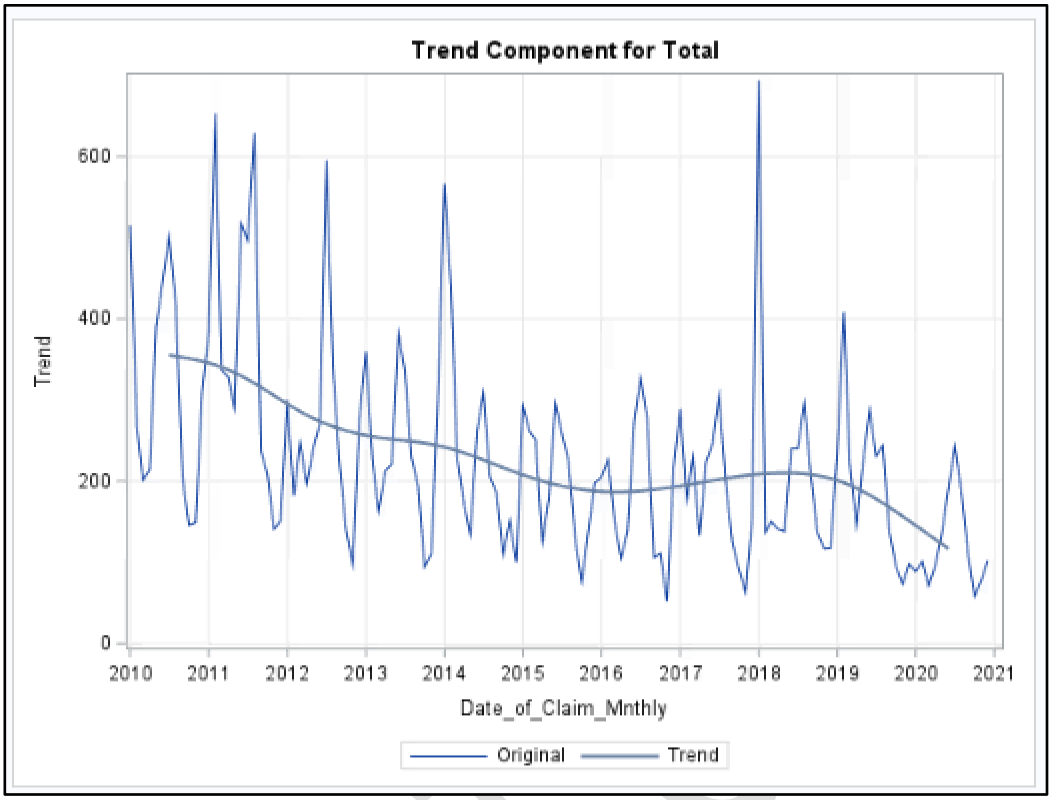
(2) Weather Model:

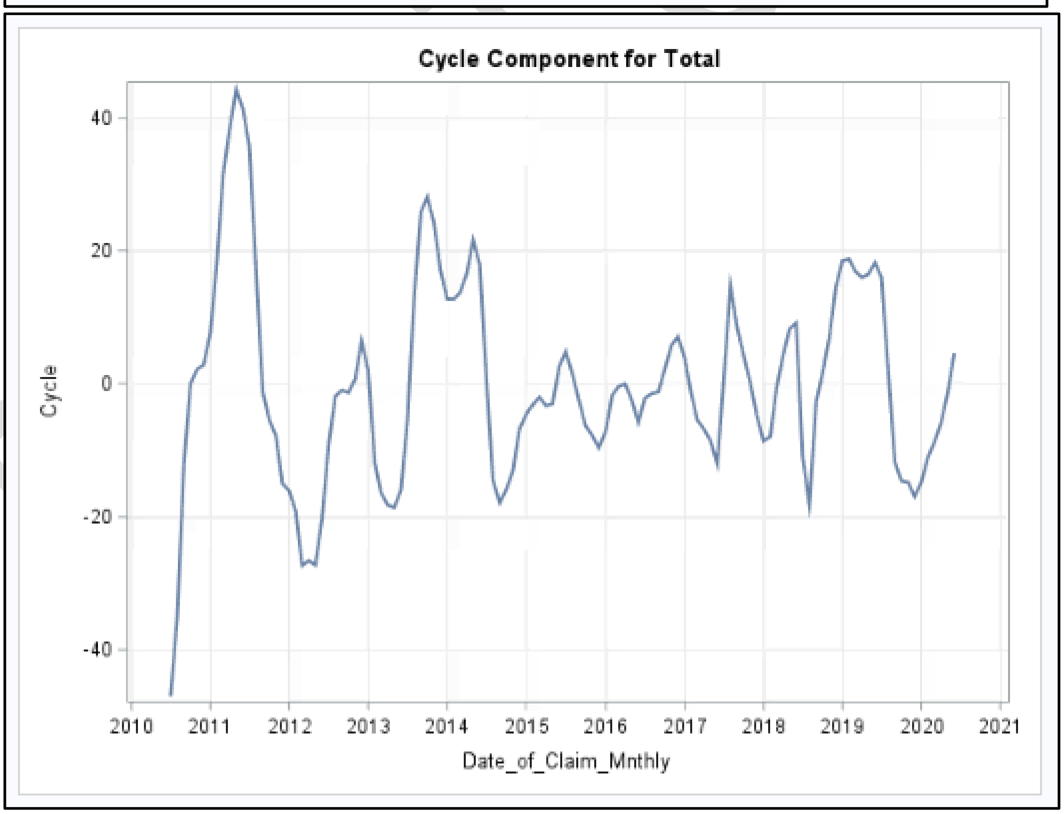
To generate the data needed for a Non-CAT weather model (excluding hail) the team used the forecasting task called “Time Series Data Preparation”. Within this task we used the following filter: CAT_IND = 'N' and Loss_Group not in (3,7) and Weather_ID = 'Y'. The task was set up to use “Total” (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and the transformation of “Total” to accumulation by means of summation.

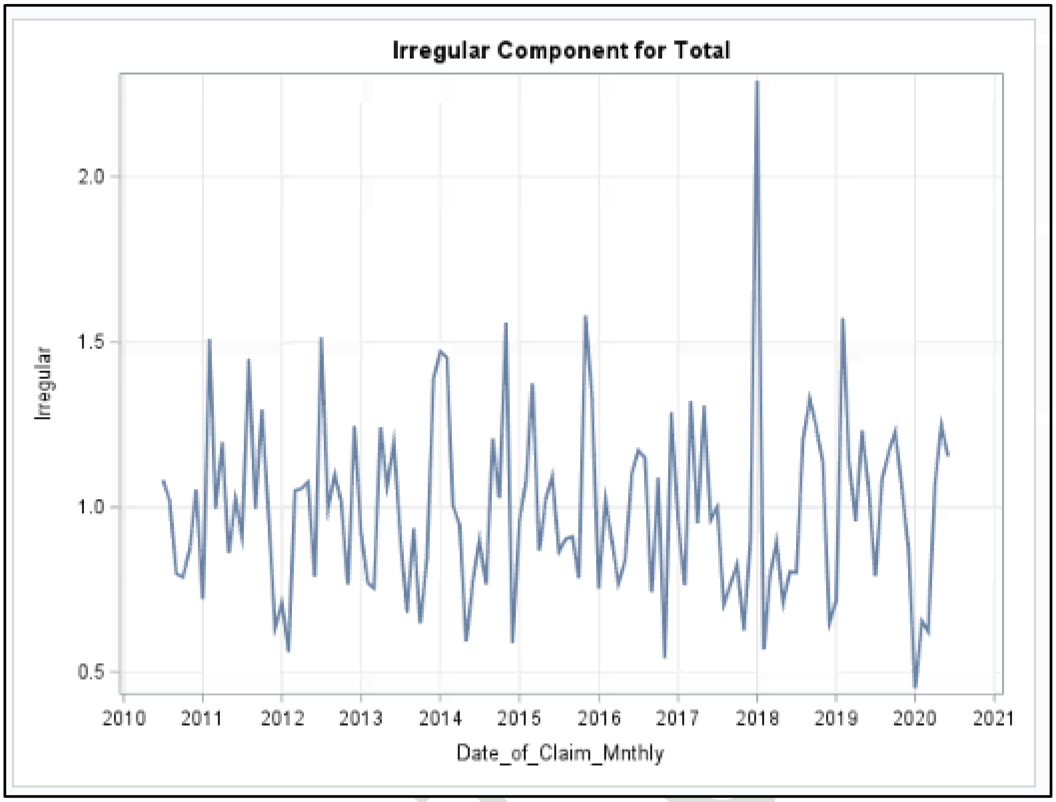
To look at the data for the Non-CAT weather model (excluding hail) model we used the forecasting task called “Time Series Exploration”. The task was set up to use “Total” (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and no transformation as the prior step took care of that. We ran a gambit of plots and reports.

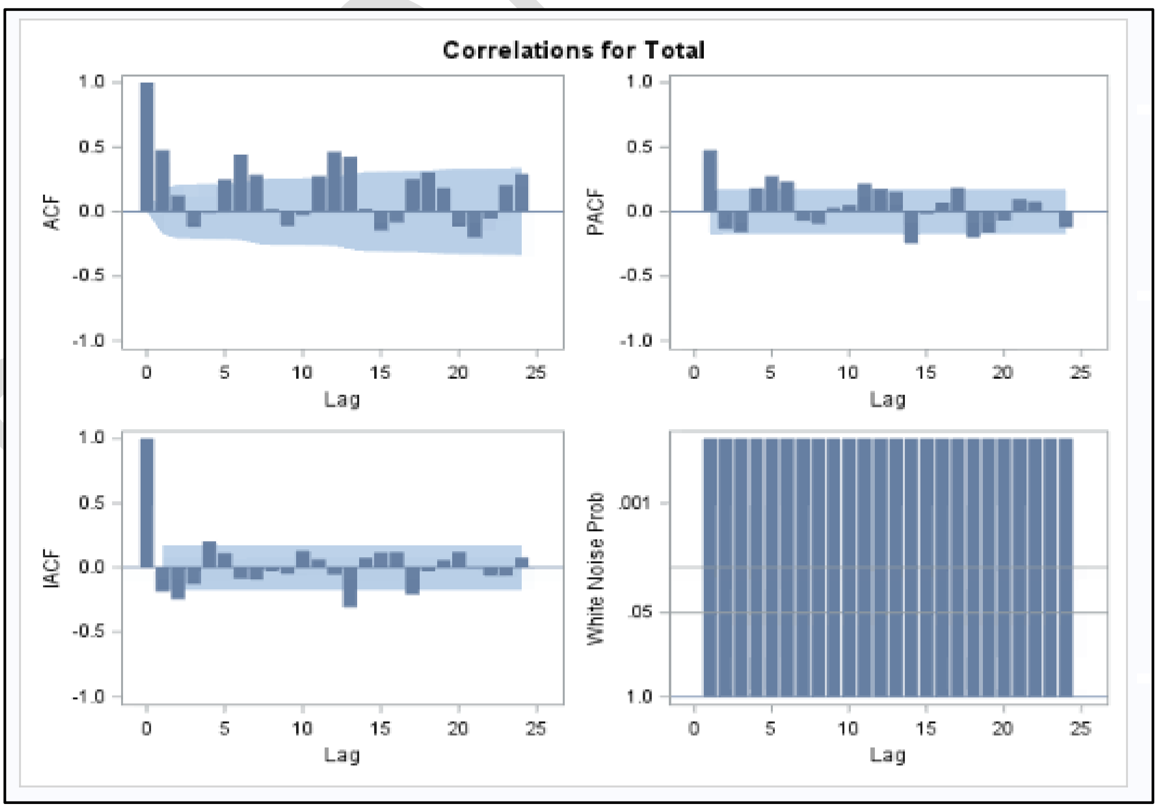
When reviewing this data set one thing was apparent and it made sense, water and wind effects are seasonal and it happened in the months of July (high) and November (low), see the season plot below. In addition, a trend has also been observed over the years where a decreasing pattern can be observed from the graph. It is interesting as the company has grown their book of business overtime. Non seasonal cycles do not appear to be an issue nor are their signals form the irregular component. Non-CAT weather (excluding hail) does have autocorrelation and is not white noise. The seasonal patterns can be observed in the ACF as well. See plots below.




Trend Component for Total
Trend
600
400
200
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly
Original
Trend


Cycle Component for Total
Cycle
40
20
0
-20
-40
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly


Irregular Component for Total
Irregular
2.0
1.5
1.0
0.5
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021
Date_of_Claim_Mnthly

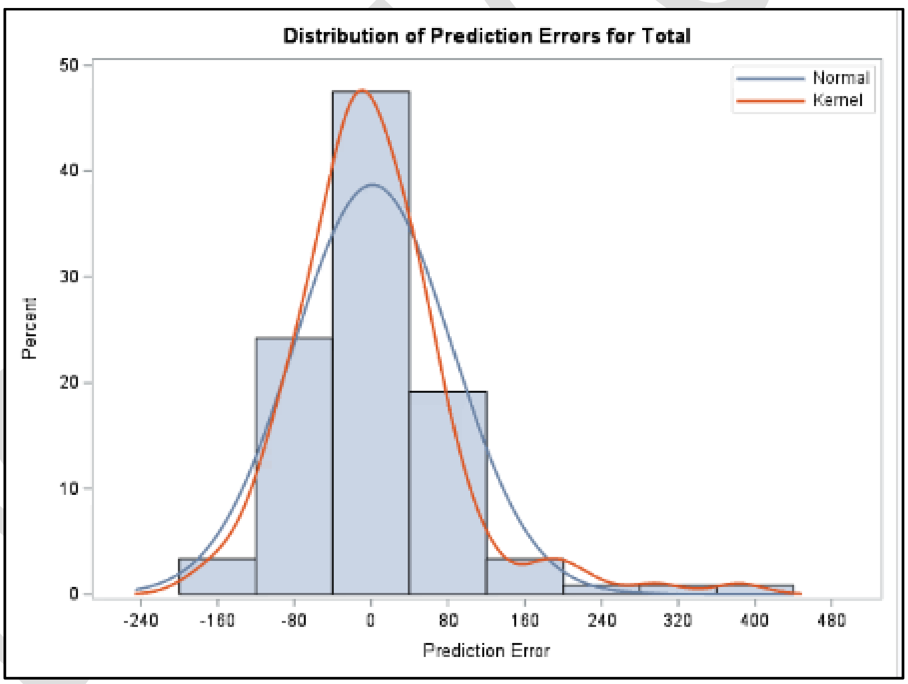

Correlations for Total
ACF
PACF
IACF
White Noise Prob
Lag
1.0 0.5 0.0 -0.5 -1.0
.001 .05 1.0
0 5 10 15 20 25

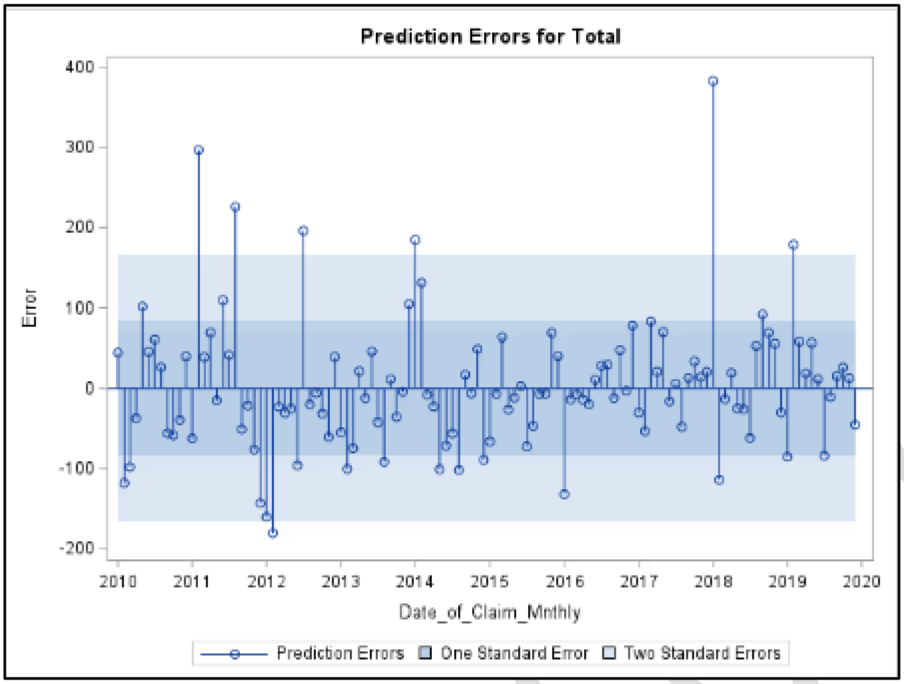
Building the Weather Model:

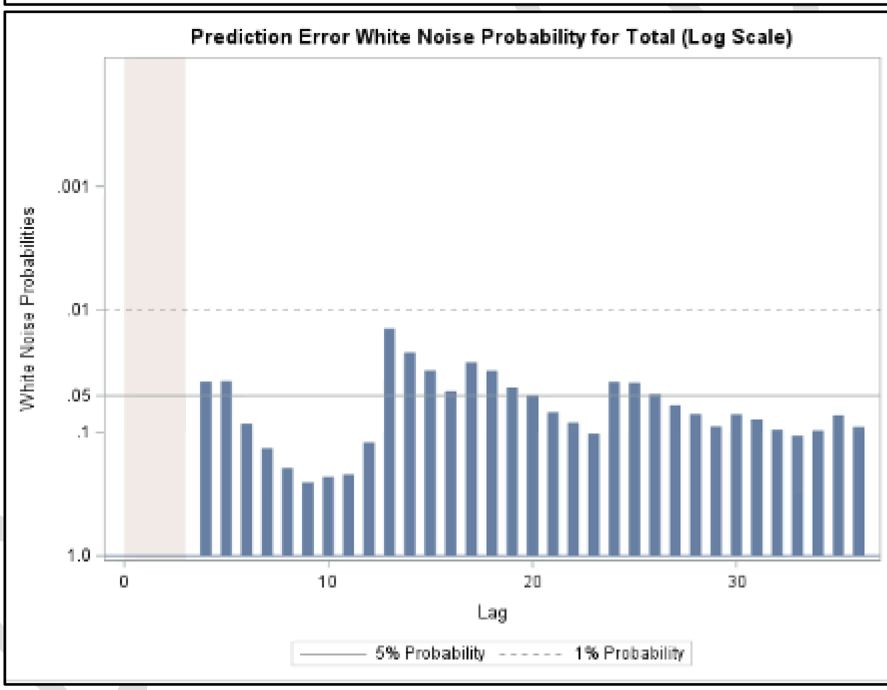
Given there is seasonality and trend the first 2 models the team explored are (1) Exponential Smoothing Model (ESM) Winter's additive and (2) ESM Winter's multiplicative models.

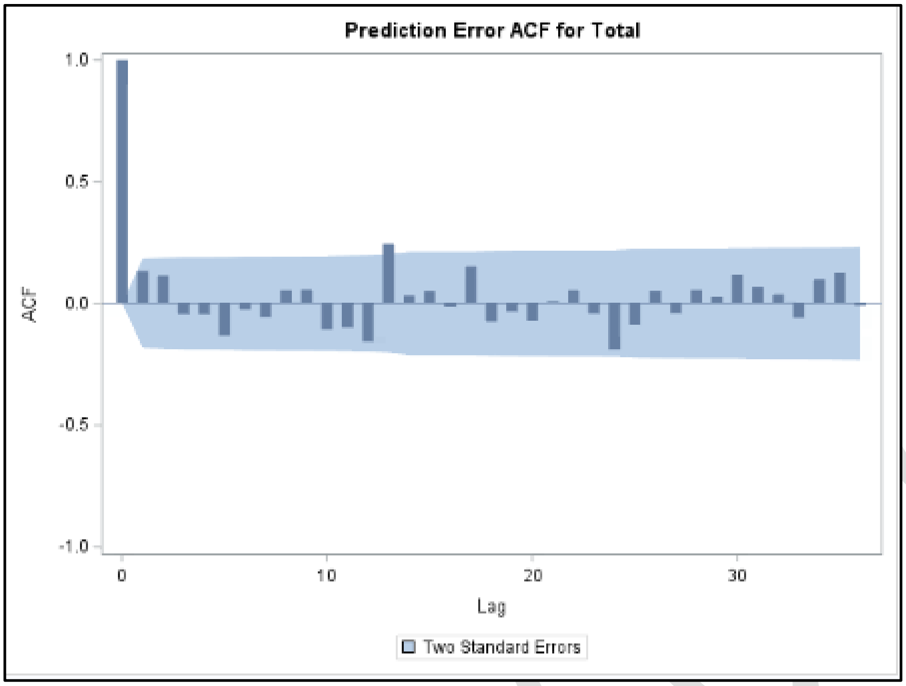
Setting up the ESM Winter's additive seasonal exponential smoothing exercise, the team started with the forecasting task "Modeling and Forecasting", selecting the PROJECT_MNTH_NCAT_NHLOTR_YWTHR table as the data. "Total" as the dependent variable, date_of_claim_mnthly as the time element. For the model, selected the ESM and selected Winter's additive exponential smoothing and ran all plots. On the options tab, select 12 months to forecast as our data is monthly and 12 months for the hold back. For the output, create fit statistics was checked and called it "waddstat.sas".

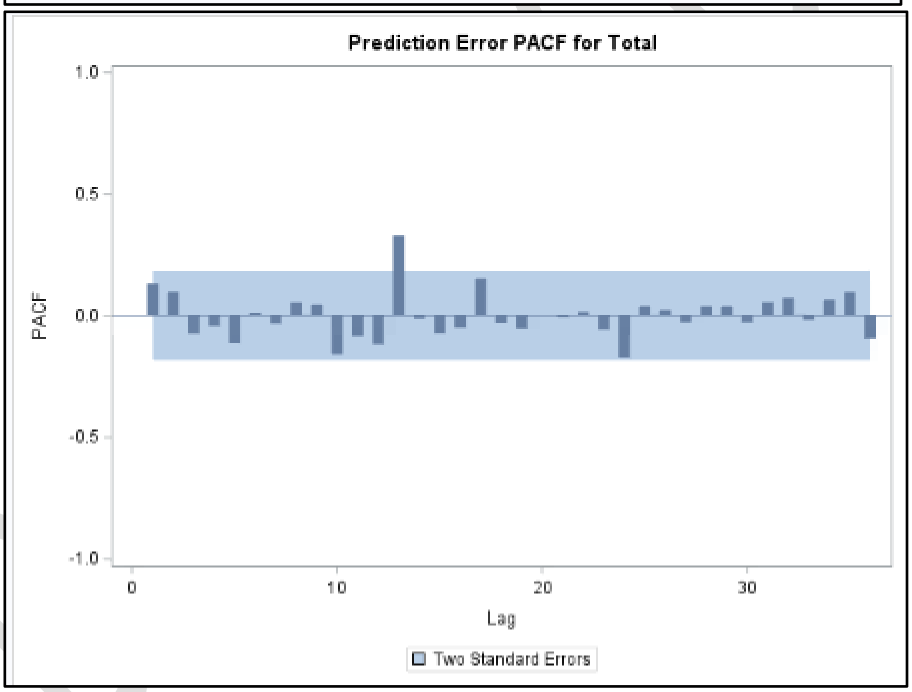
Overall the model did well, errors were normally distributed, mostly within 1 and 2 standard deviations and were White Noise. The ACF and PACF on the errors do not show signs of autocorrelation, with a few lags on the IACF being somewhat significant. When eyeballing the forecast plot the model seems to do a reasonable job at following the seasonality and trend. See plots below. The AIC on the fit was 1,064.11 (rounded) and the SBC was 1,072.47 (rounded). The MAPE and RMSE on the fit were 25.10 (rounded) and 82.17 (rounded), respectively. The AIC on the forecast was 106.57 (rounded) and the SBC was 106.57 (rounded). The MAPE and RMSE on the forecast were 56.86 (rounded) and 84.80 (rounded), respectively.

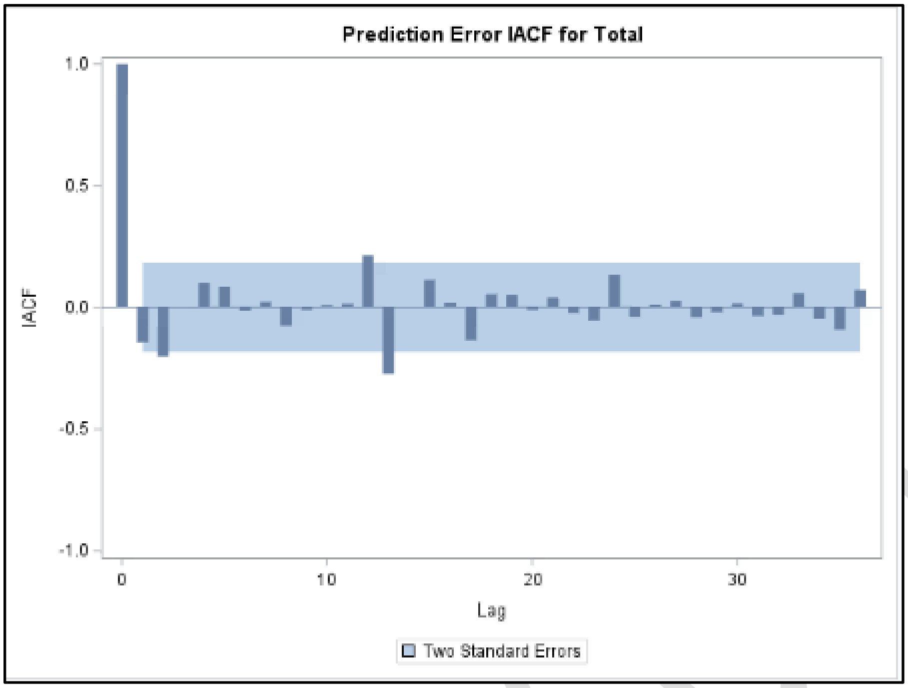


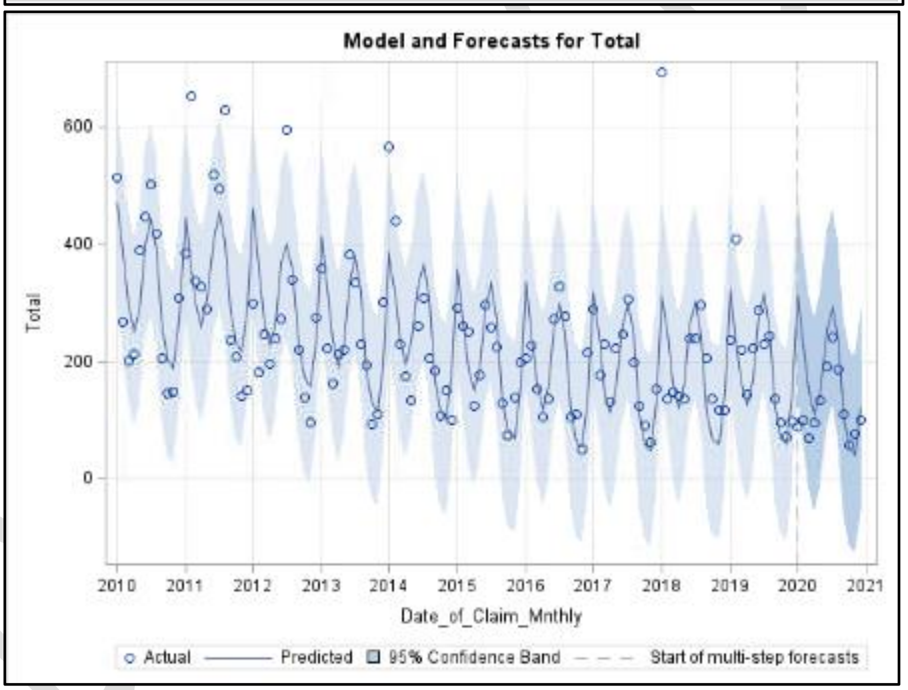

Prediction Errors for Total
Error
400
300
200
100
0
-100
-200
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Date_of_Claim_Mnthly
Prediction Errors
One Standard Error
Two Standard Errors


Prediction Error White Noise Probability for Total (Log Scale)
White Noise Probabilities
.001
.01
.05
.1
1.0
0
10
20
30
Lag
5% Probability
1% Probability


Prediction Error ACF for Total
ACF
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
Lag
Two Standard Errors

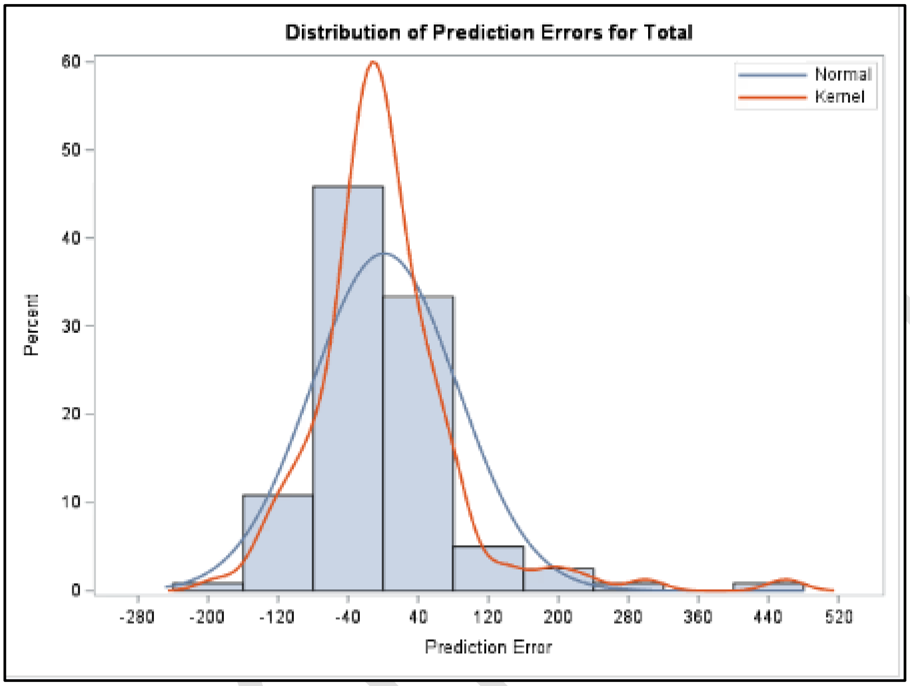

Prediction Error PACF for Total
PACF
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
Lag
Two Standard Errors

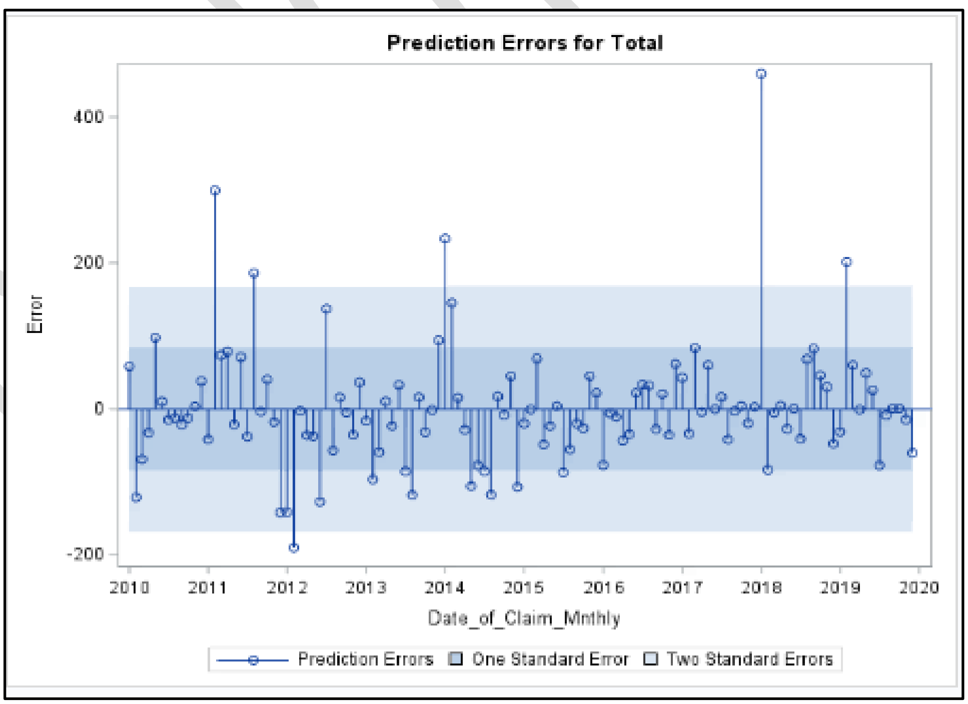


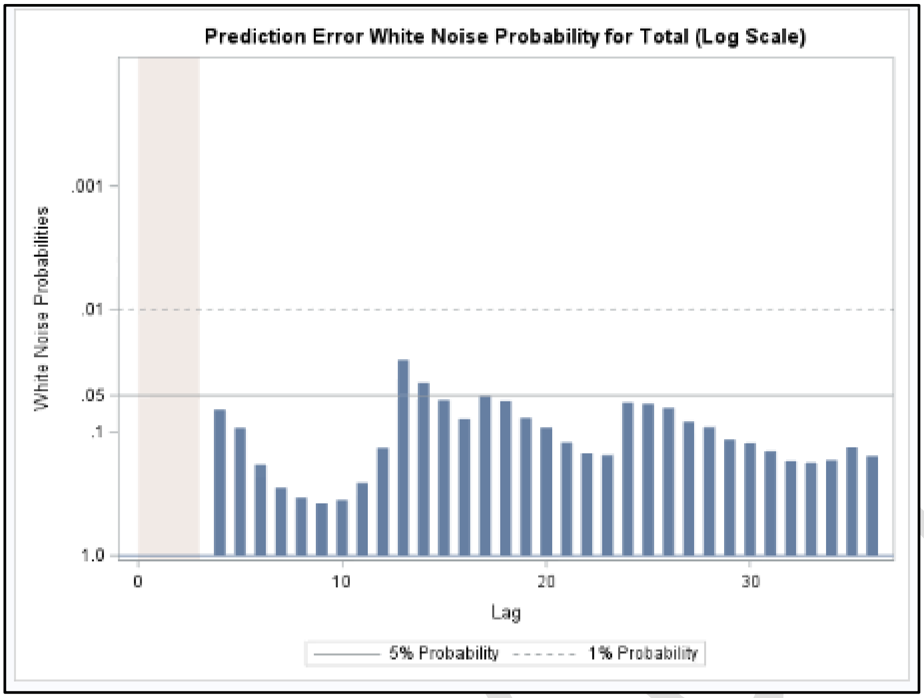


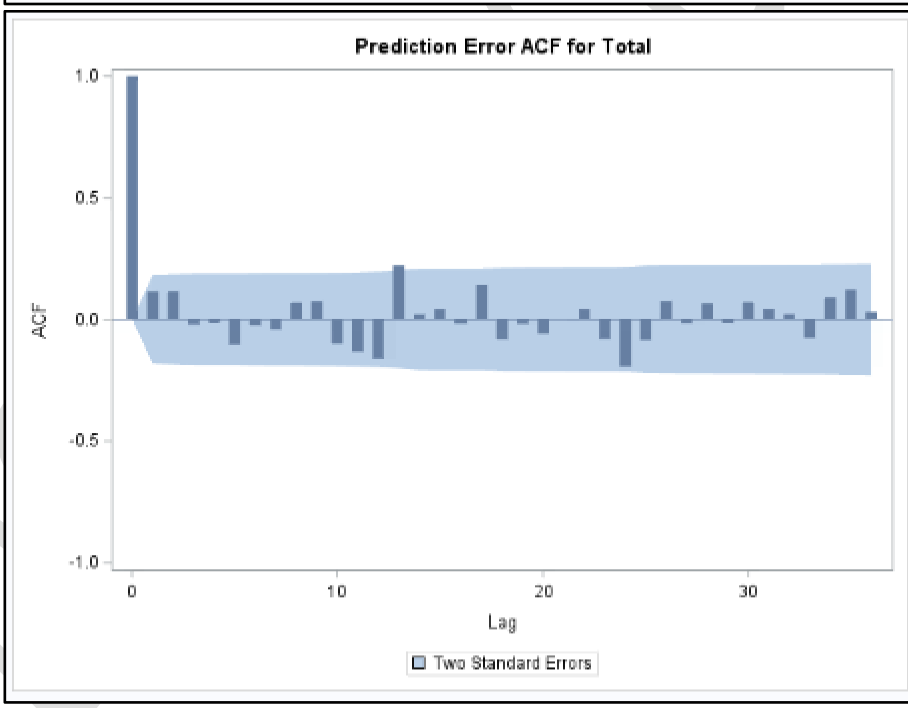
Setting up the ESM Winter's multiplicative exercise, the team started with the forecasting task "Modeling and Forecasting", selecting the PROJECT_MNTH_NCAT_NHLOTR_YWTHR table as the data. "Total" as the dependent variable, date_of_claim_mnthly as the time element. For the model, selected the ESM and selected Winter's multiplicative and ran all plots. On the options tab, select 12 months to forecast as our data is monthly and 12 months for the hold back. For the output, create fit statics was checked and called it "wmulstat.csv".

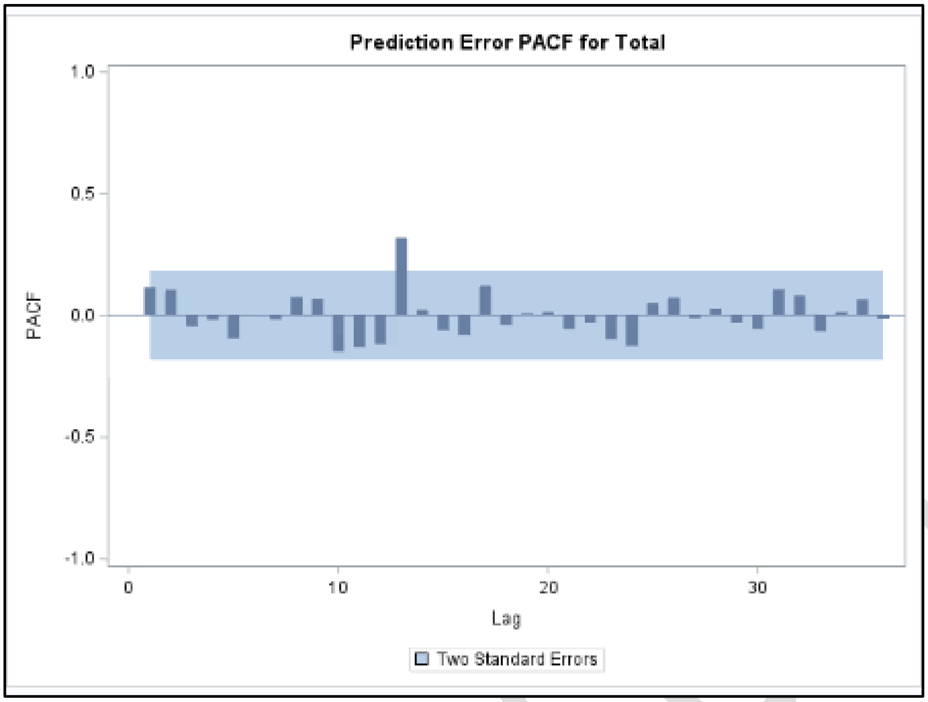
Overall the model did well, errors were normally distributed, mostly within 1 and 2 standard deviations and were White Noise. The ACF and PACF on the errors do not show signs of autocorrelation, with a few lags on the IACF being somewhat significant. When eyeballing the forecast plot the model seems to do a reasonable job at following the seasonality and trend. See plots below. The AIC on the fit was 1,066.85 (rounded) and the SBC was 1,075.22 (rounded). The MAPE and RMSE on the fit were 23.03 (rounded) and 83.11 (rounded), respectively. The AIC on the forecast was 100.43 (rounded) and the SBC was 100.43 (rounded). The MAPE and RMSE on the forecast were 49.10 (rounded) and 65.66 (rounded), respectively.

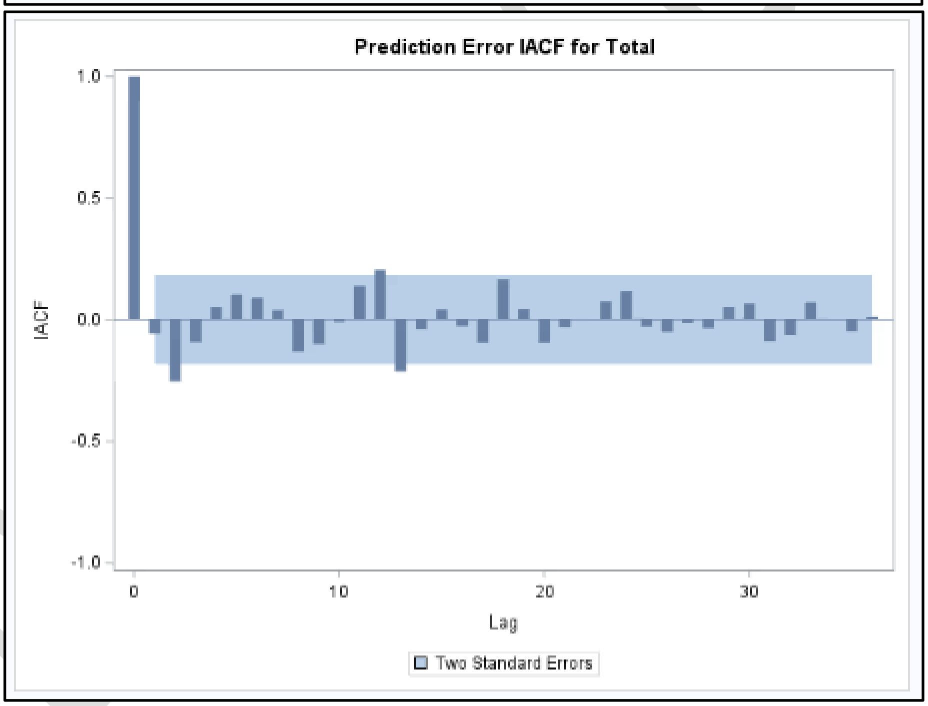


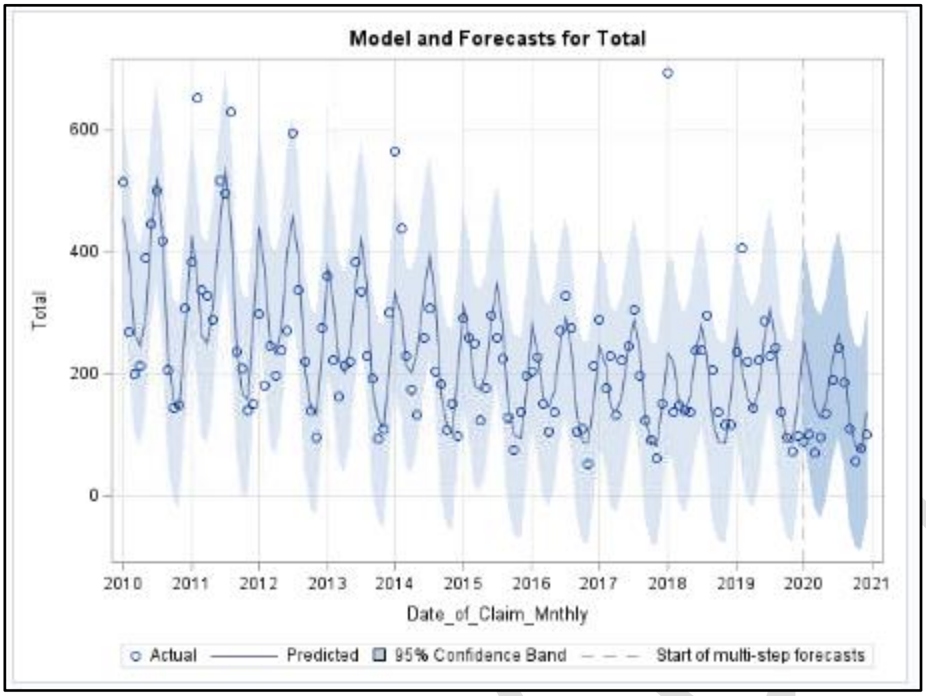



Prediction Error White Noise Probability for Total (Log Scale)
White Noise Probabilities
.001
.01
.05
.1
1.0
0
10
20
30
Lag
5% Probability
1% Probability


Prediction Error ACF for Total
ACF
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
Lag
Two Standard Errors


Prediction Error PACF for Total
PACF
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
Lag
Two Standard Errors


Prediction Error IACF for Total
IACF
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
Lag
Two Standard Errors



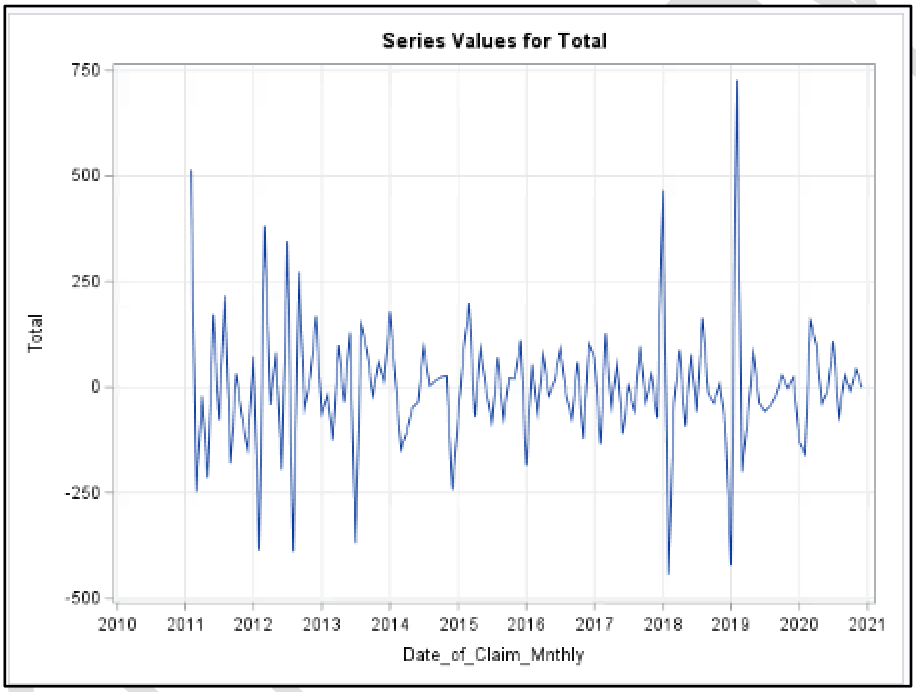
Between the 2 models the statistical metrics are slightly higher for the Winter's Additive model. However, **ESM Winter's Multiplicative** did a better job with respect to the white noise. Therefore, ESM Winter's Multiplicative will be the model to be chosen for the Weather model.

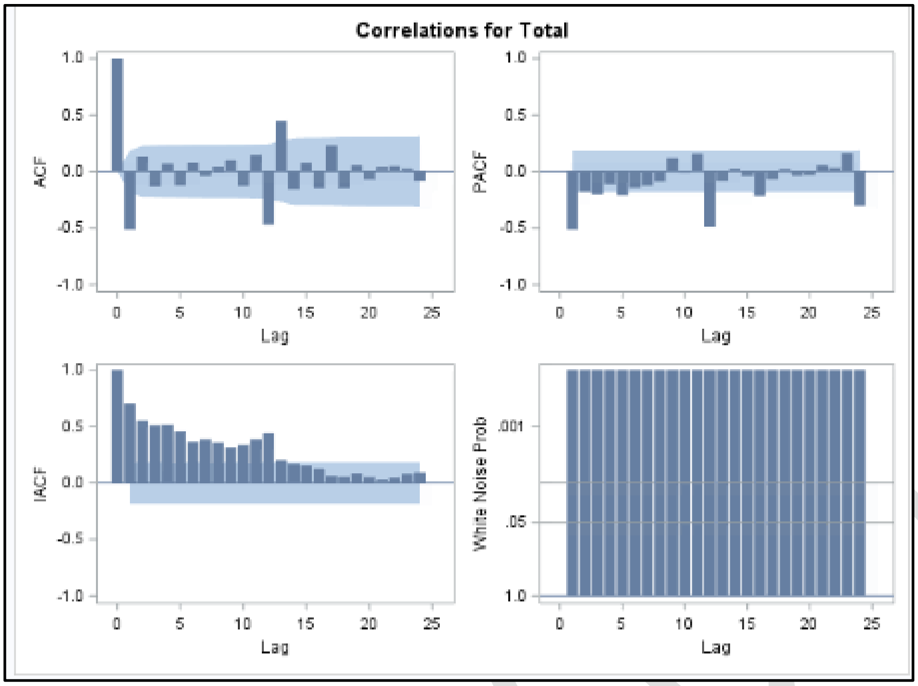
ARIMA models were considered however they did not make the final model selection process because we could not get comfortable the significance levels when selecting the different moving average order level. In addition the autocorrelation plots looked decent but the white noise plots still show significance at different lags. We tried multiple different combinations of autoregressive and moving average assumptions. In addition, for each combination we tried, the SBC and the AIC remained over 1,400 when compared to the ESM models, which were around 1,000 - 1,100. Ultimately, we were not going to choose the ARIMA model. Below is an example for the most successful combination. The data had to be transformed with 1 simple differencing and 1 seasonal differencing because 1 seasonal differencing did a good job getting rid of seasonality but the trend remained. Though 1 seasonal differencing took care of the trend, adding a 1 simple differencing removed the trend.

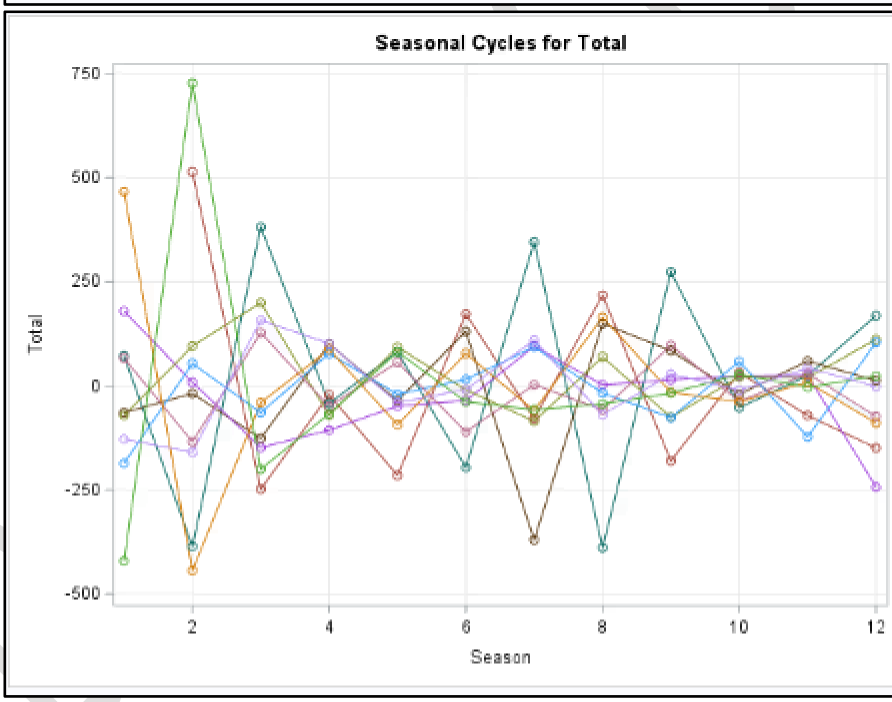
To generate the data needed for an ARIMA model the team used the forecasting task called "Time Series Data Preparation". Within this task we leveraged the table PROJECT_MNTH_NCAT_NHLOTR_YWTHR. The task was set up to use "Total" (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and the transformation of "Total" to accumulation by means of summation, simple differencing = 1 and seasonal differencing = 1.

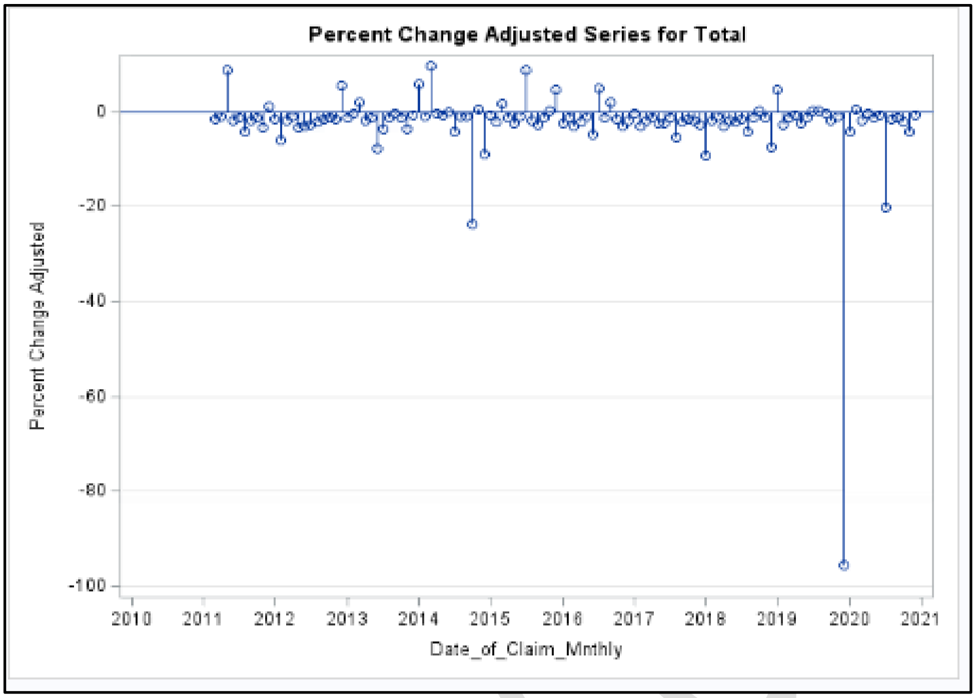
To look at the result of 1 simple differencing & 1 seasonal differencing on the Weather data we used the forecasting task called “Time Series Exploration ". The task was set up: “Total” (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly. We ran a gambit of plots and reports.

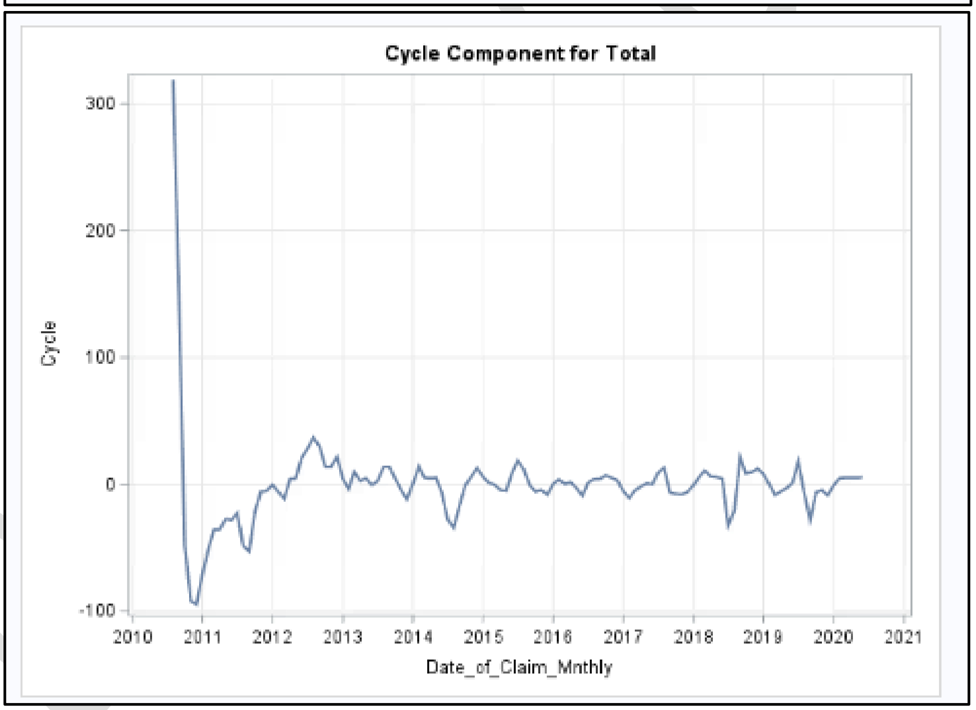
It can be seen that the differencing applied to the data took care of the trend which has essentially been eliminated. It also took care of the seasonality, which has been reduced and nearly eliminated. There still remains no cycle or irregular component. However, the correlation plots do still show autocorrelation and the data is not white noise, meaning a model could still be made to capture that remaining signal. We focused on autoregressive <=1 because the IACF and the PACF had significant spikes in lag 1 and are gradually reducing and we looked at the moving average <=1 because the ACF has a significant spike in lag 1. See plots below.

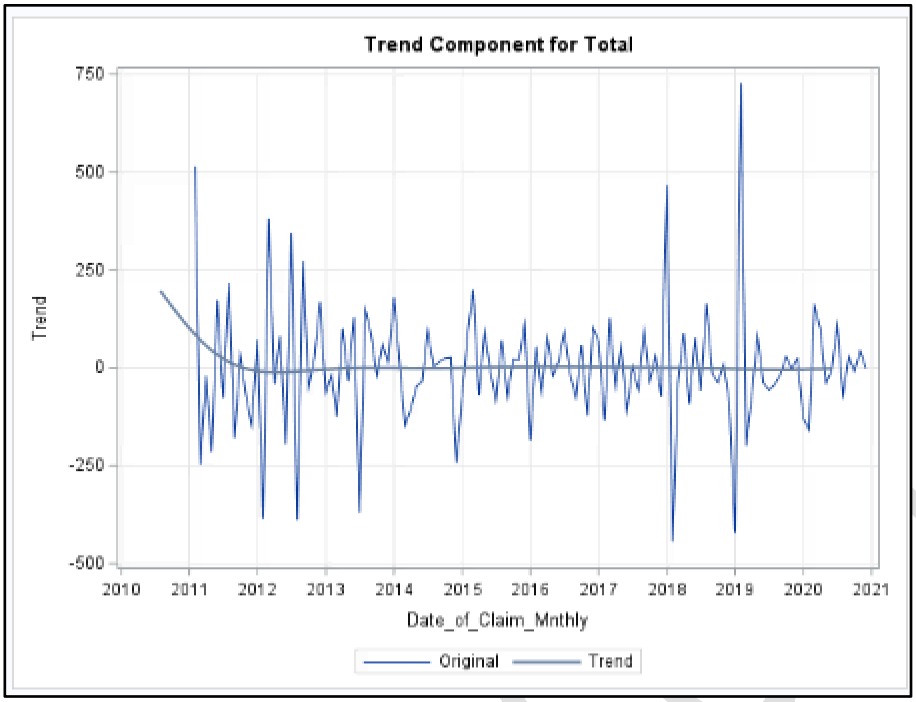


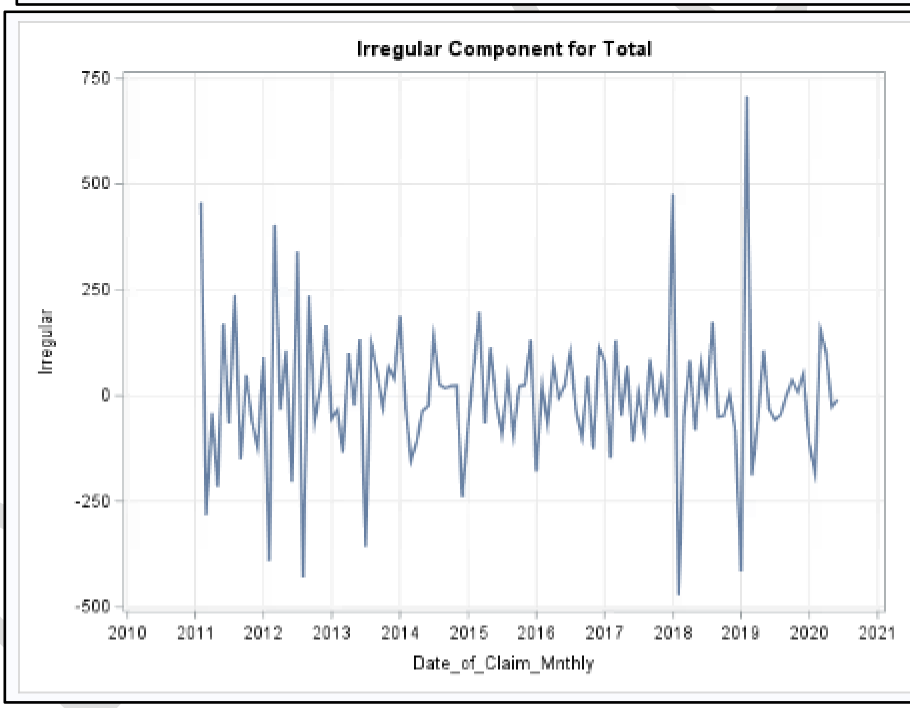

Correlations for Total
ACF
PACF
IACF
White Noise Prob
Lag


Seasonal Cycles for Total
Total
Season


Percent Change Adjusted Series for Total
Percent Change Adjusted
0
-20
-40
-60
-80
-100
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly


Cycle Component for Total
Cycle
300
200
100
0
-100
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly





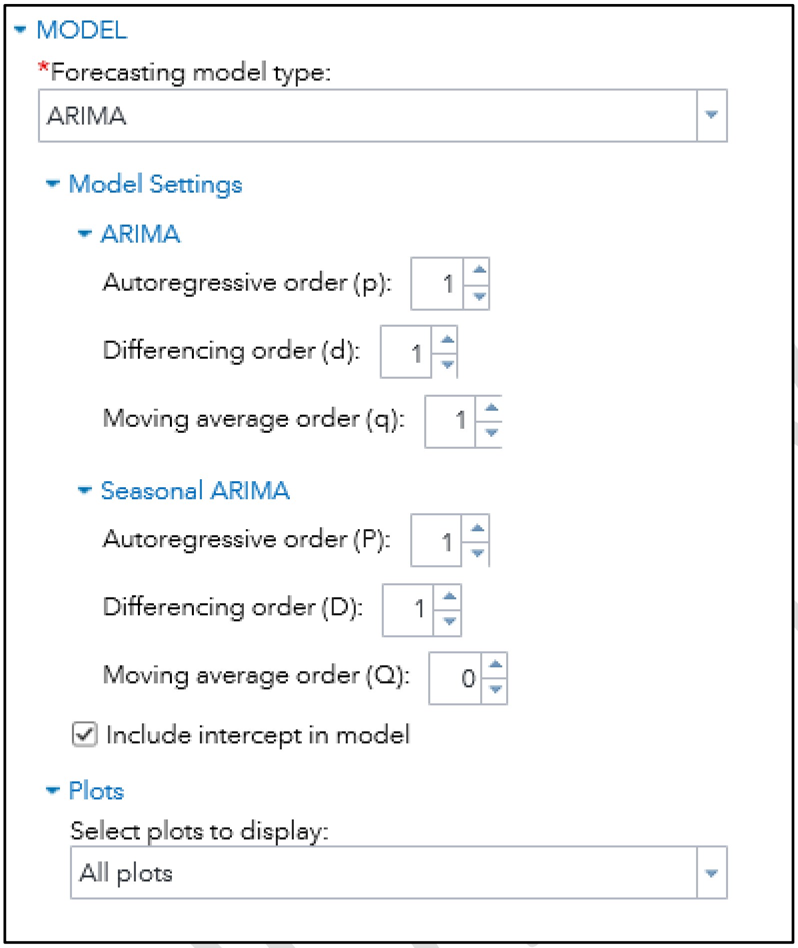
Setting up the best ARIMA model, the team started with the forecasting task "Modeling and Forecasting", selecting the PROJECT_MNTH_NCAT_NHLOTR_YWTHR_D table as the data. "Total" as the dependent variable, date_of_claim_mnthly as the time element. For the model, selected the ARIMA with and selected p=1 d=1, q=1, P=1, D=1, Q=0 and ran all plots. On the options tab, select 12 months to forecast as our data is monthly and 12 months for the hold back.

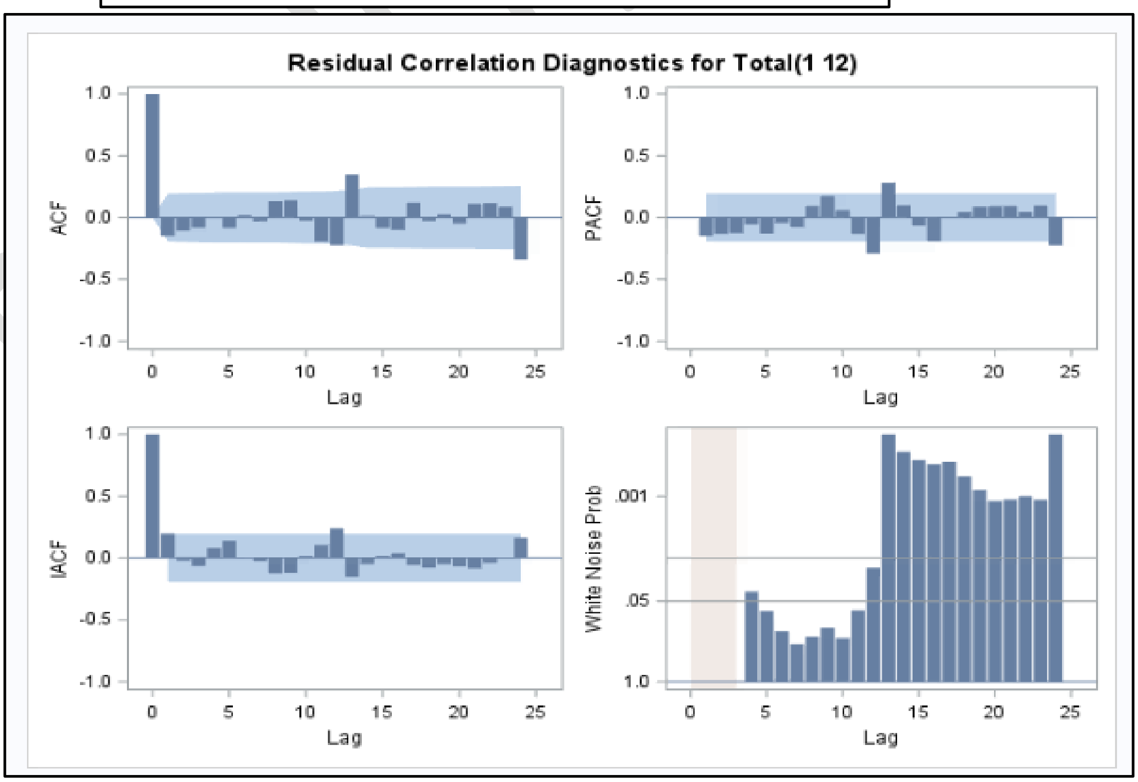
Originally the thought was this was going to be a moving average selection because the PACF and the IACF are gradually decaying from lag 0 but through our trial and error the result showed that the model we built is disliking the moving average selections because it kept failing the p-test were significant. Below is the Maximum likelihood estimation for our best effort:

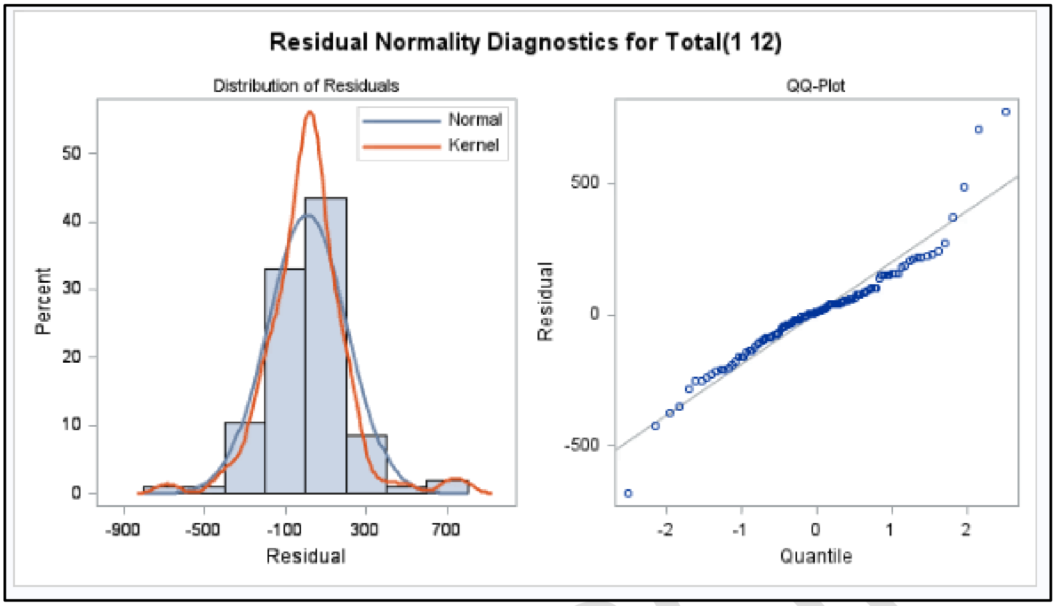
ARIMA Estimation Optimization Summary	
Estimation Method	Maximum Likelihood
Parameters Estimated	4
Termination Criteria	Maximum Relative Change in Estimates
Iteration Stopping Value	0.001
Criteria Value	73.46906
Maximum Absolute Value of Gradient	1029922
R-Square Change from Last Iteration	0.516116
Objective Function	Log Gaussian Likelihood
Objective Function Value	-714.84
Marquardt's Lambda Coefficient	0.00001
Numerical Derivative Perturbation Delta	0.001
Iterations	11
Warning Message	Estimates may not have converged.

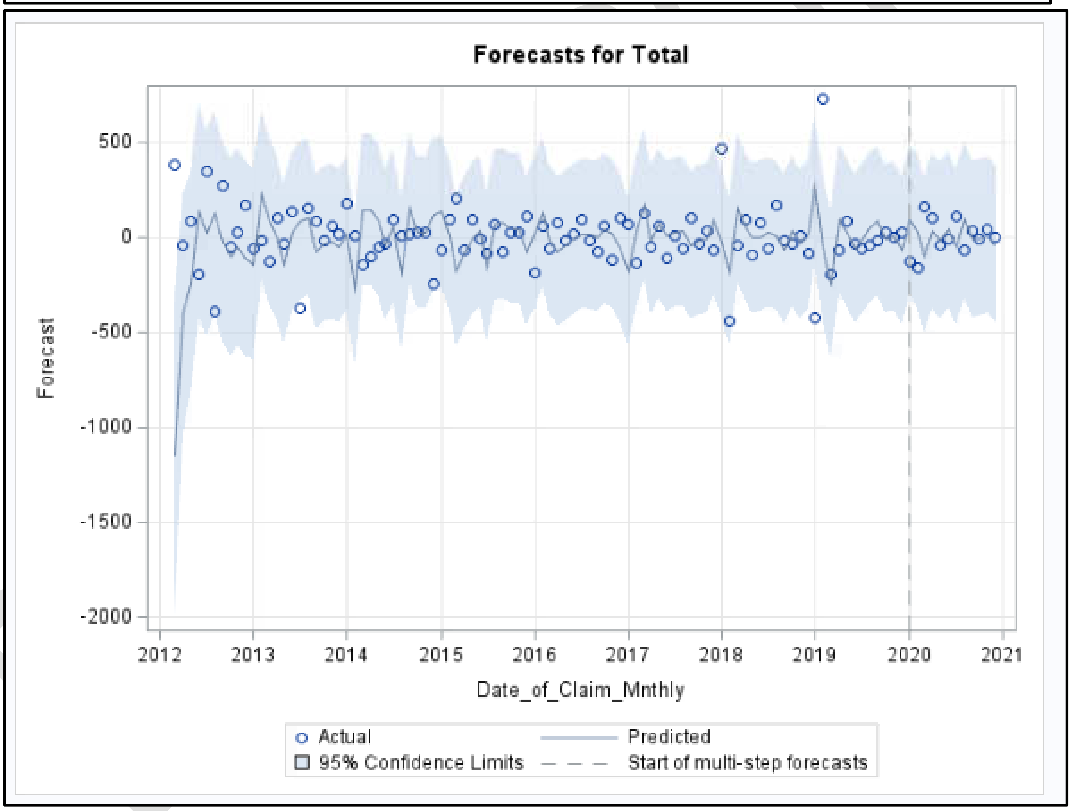
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.04811	0.43980	0.11	0.9129	0
MA1,1	0.99975	14.72966	0.07	0.9459	1
AR1,1	-0.32022	0.08994	-3.56	0.0004	1
AR2,1	-0.60984	0.07346	-8.30	<.0001	12

You can see that the model did not like MA1,1 with a 0.9459 for the |t| test, however if we removed the MA selection the autocorrelation in the residuals came back and the white noise test looked worse with more significant bars. Below are the results of the best options we picked with the lowest SBC and AIC of 1,448.33 and 1,437.68, respectively. The residuals are normally distributed, the ACF/PACF/IACF for the residuals look reasonable and not significantly but the white noise is elevated. This would have been a contenting model if we felt better about the t-test for MA but since it was outside our comfort level and more importantly the SBC could not match the ESM models we did not pursue future refinement. Below the plots, we provided an example of removing the MA selection.

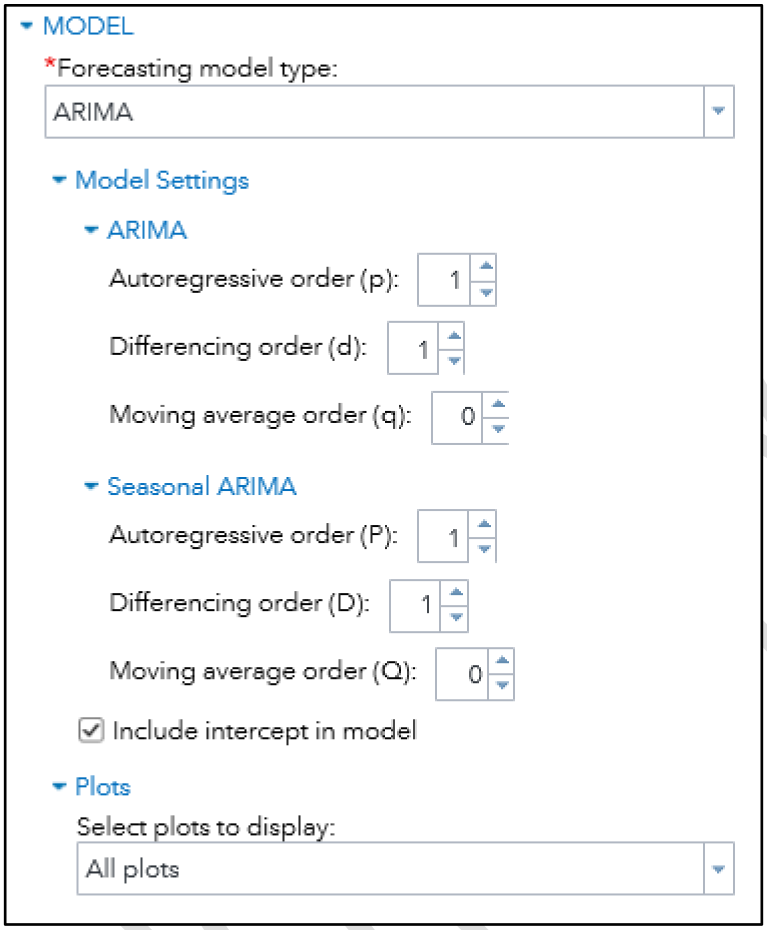

MODEL
*Forecasting model type:
ARIMA
Model Settings
ARIMA
Autoregressive order (p): 1
Differencing order (d): 1
Moving average order (q): 1
Seasonal ARIMA
Autoregressive order (P): 1
Differencing order (D): 1
Moving average order (Q): 0
Include intercept in model
Plots
Select plots to display:
All plots

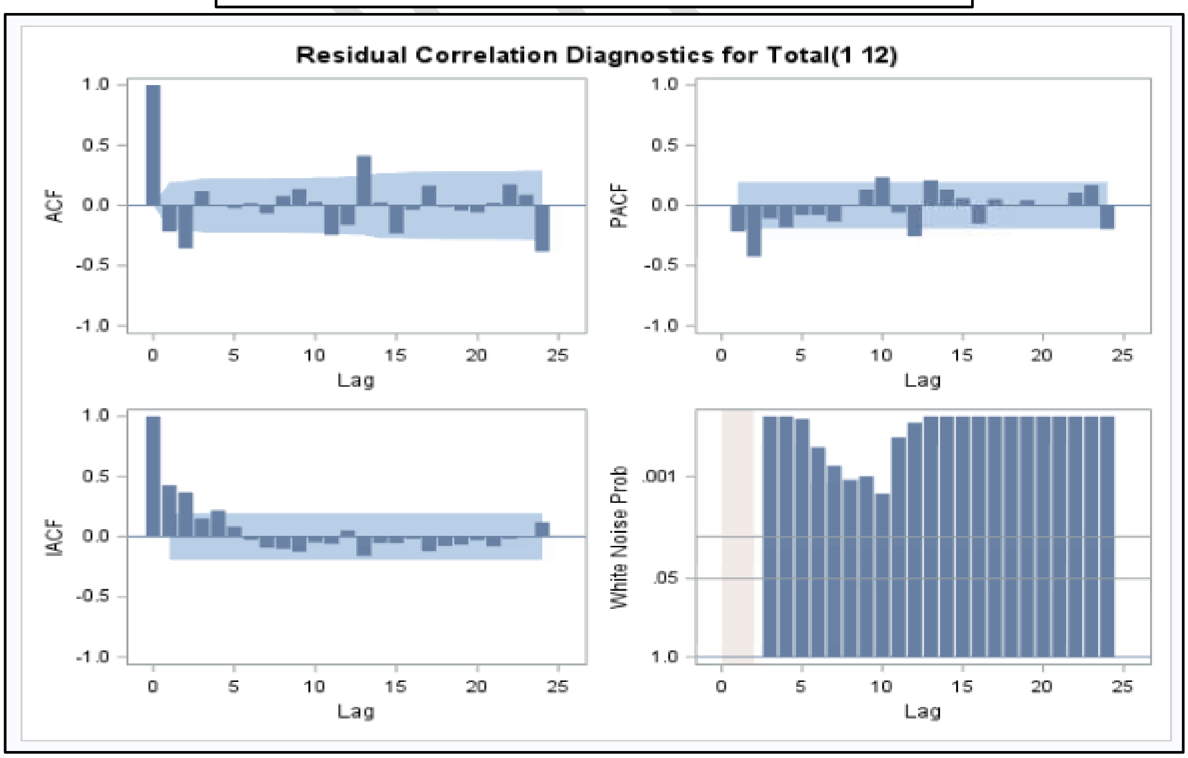

Residual Correlation Diagnostics for Total(1 12)
ACF
PACF
IACF
White Noise Prob
Lag
1.0
0.5
0.0
-0.5
-1.0
.001
.05
0
5
10
15
20
25

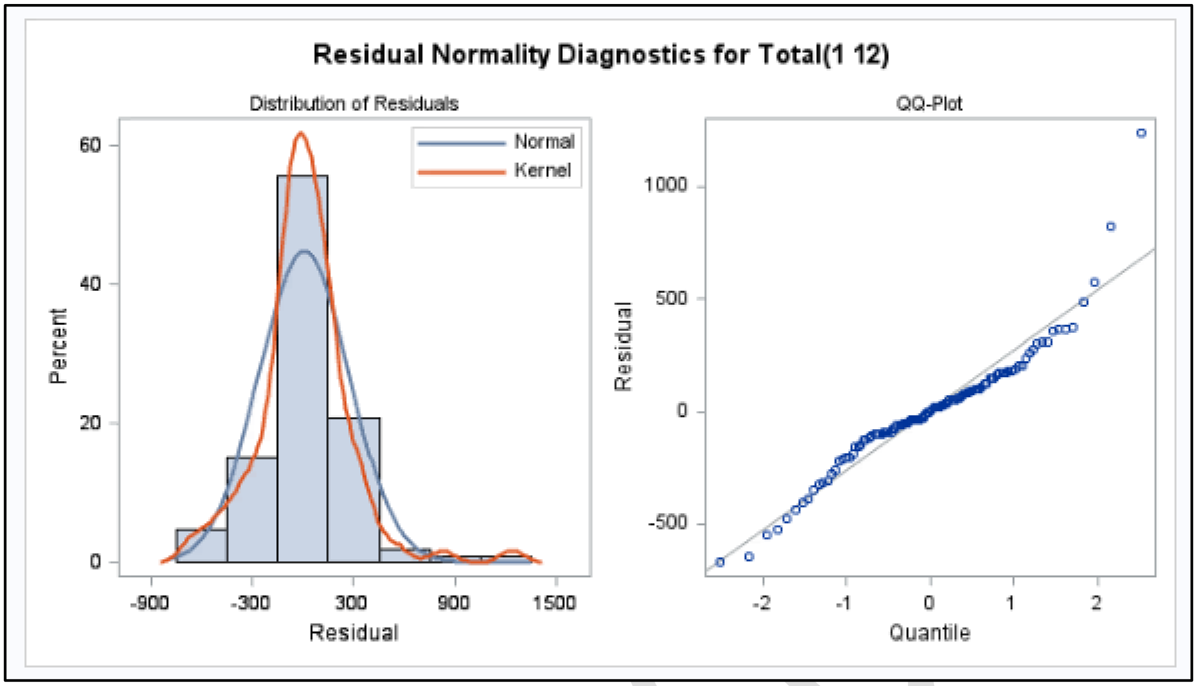


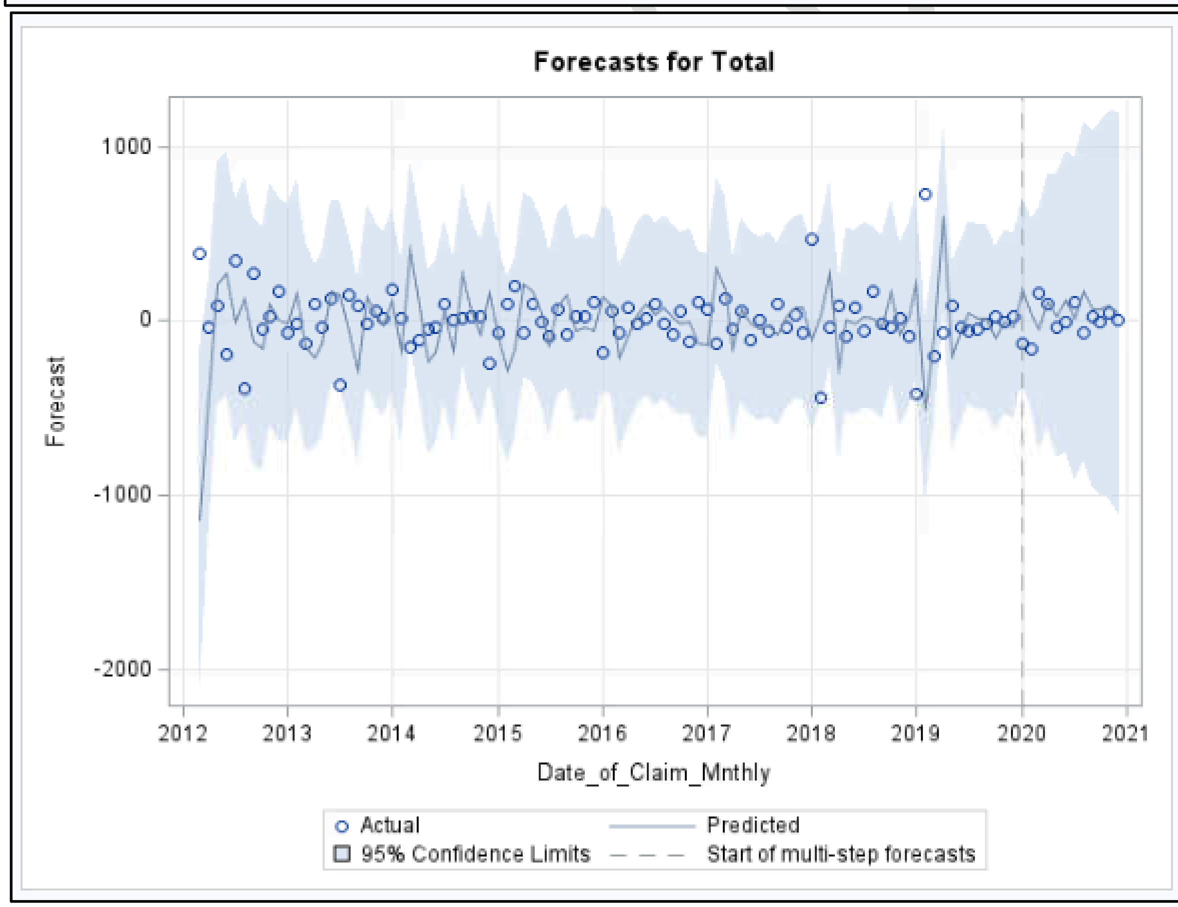


Example of removing the MA selection from our best attempt. The SBC and AIC deteriorate (1,506.01 and 1,498.02 respectively) but the AR selections both p and P show they are not significant and pass the t-test. However the PACF and IACF are showing autocorrelation and the whitenoise test looks terrible as all the lags are significant.


MODEL
*Forecasting model type:
ARIMA
Model Settings
ARIMA
Autoregressive order (p): 1
Differencing order (d): 1
Moving average order (q): 0
Seasonal ARIMA
Autoregressive order (P): 1
Differencing order (D): 1
Moving average order (Q): 0
Include intercept in model
Plots
Select plots to display:
All plots


Residual Correlation Diagnostics for Total(1 12)
ACF
PACF
IACF
White Noise Prob
Lag


Residual Normality Diagnostics for Total(1 12)
Distribution of Residuals
QQ-Plot
Normal
Kernel
Percent
60
40
20
0
-900
-300
300
900
1500
Residual
1000
500
0
-500
-2
-1
0
1
2
Quantile

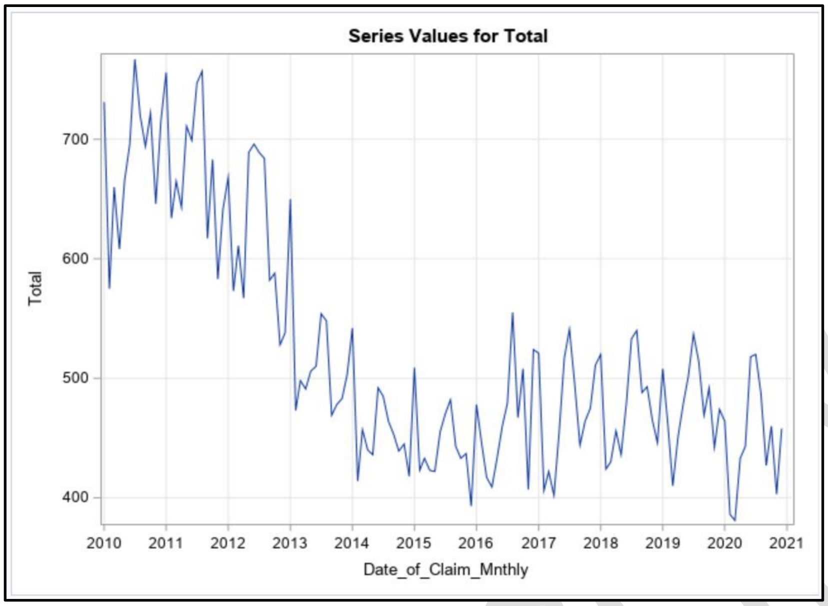

Forecasts for Total
Forecast
1000
0
-1000
-2000
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Date_of_Claim_Mnthly
Actual
Predicted
95% Confidence Limits
Start of multi-step forecasts

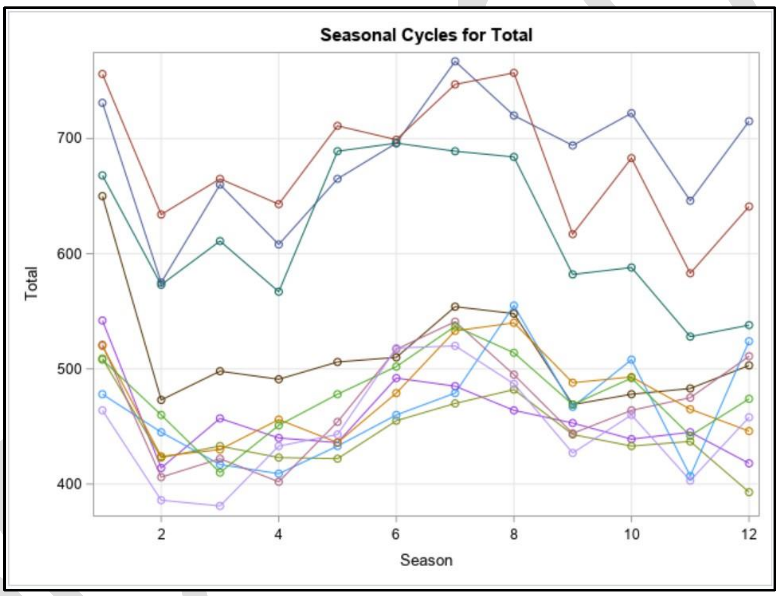
(3) xWeather Model:

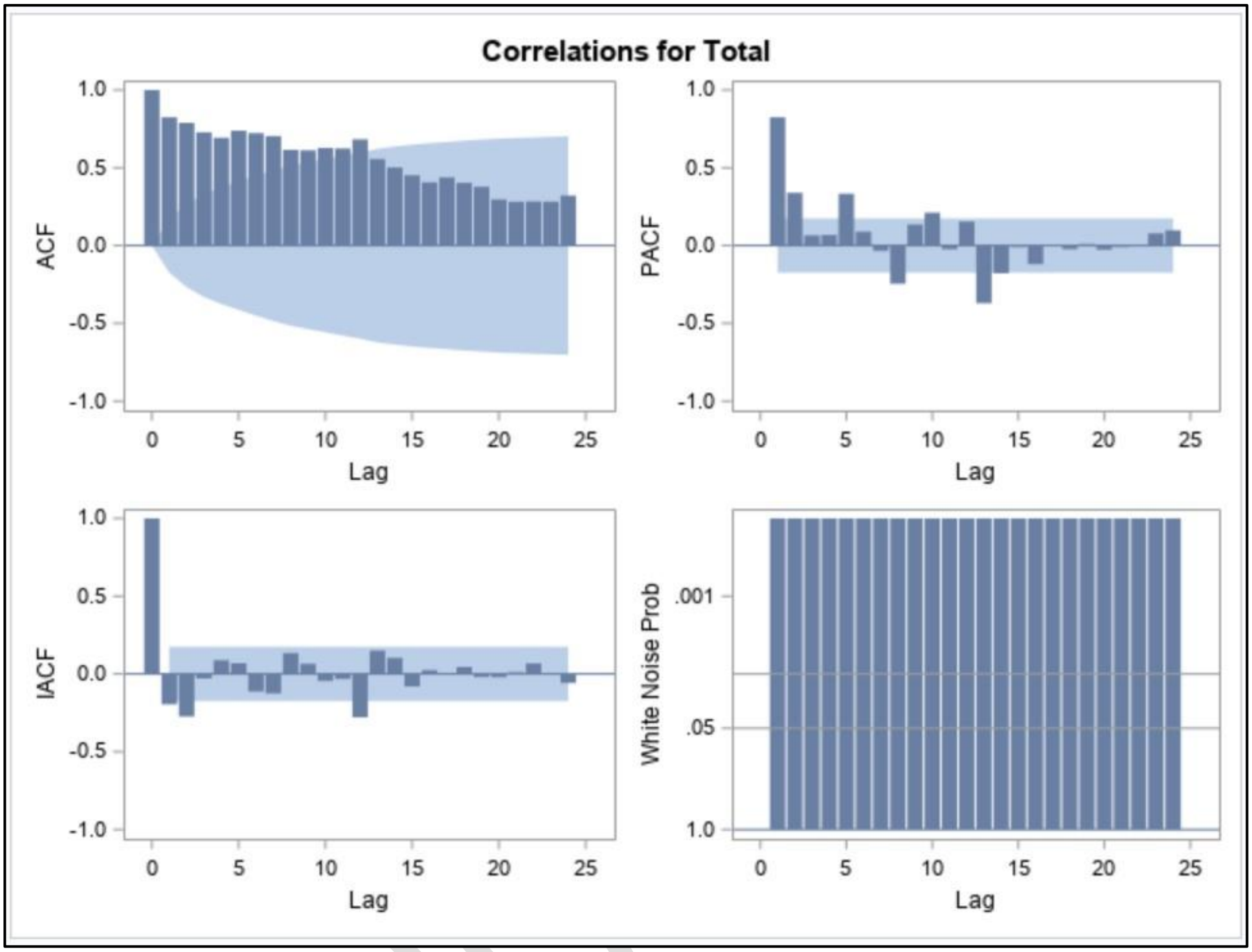
To generate the data needed for a xWeather model (excluding "All Other") the team used the forecasting task called "Time Series Data Preparation". Within this task we used the following filter: CAT_IND = 'N' and Loss_Group not in (3,7) and Weather_ID = 'N'. The task was set up to use "Total" (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and the transformation of "Total" to accumulation by means of summation.

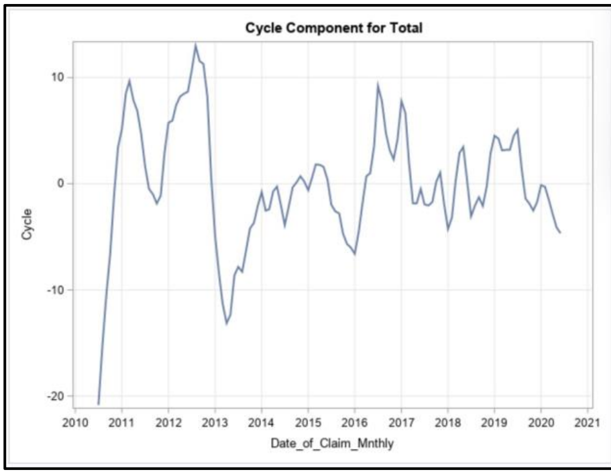
To look at the data for the xWeather model we used the forecasting task called "Time Series Exploration". The task was set up to use "Total" (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and no transformation as the prior step took care of that. We ran a gambit of plots and reports.

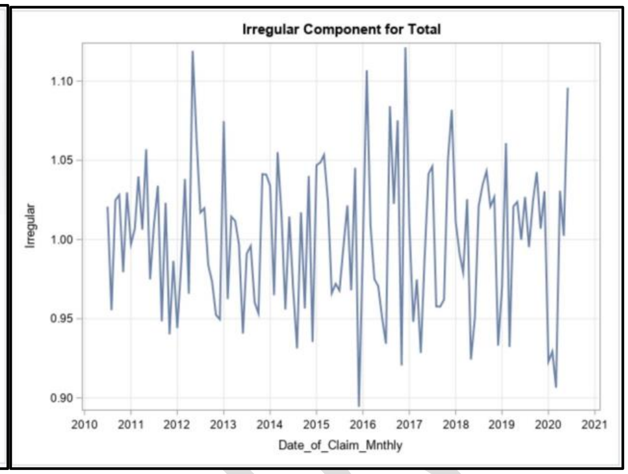
When reviewing this data set there were a few interesting findings worth noting. The first of these is that there is a clear trend present in the data. There is a steep drop-off in the number of these claims and then a slight rise towards the end of the series. This can be seen in the trend component plot below. Something that came as more of a surprise when looking at the data was that there seemed to be a seasonal component to this series. This can be seen in the seasonal cycles plot and the seasonal component plot. The types of claims included in this non-cat, non-weather, non-hail data would be damage such as part failure, building damage due to pipe breaks, or human error that led to damages. Originally this didn't make sense to us as this type of claim should be uniformly distributed throughout the year. However, upon further exploration into the data, it was found that the seasonal component was coming from pipe failure due to temperature fluctuations in the year. As can be seen below, the cycle and irregular components are not significant in this time series. It can also be seen that there is some autocorrelation in the correlation panel, especially in the ACF plot. There is no white noise, meaning that there is a lot of signal that can be captured in a model.

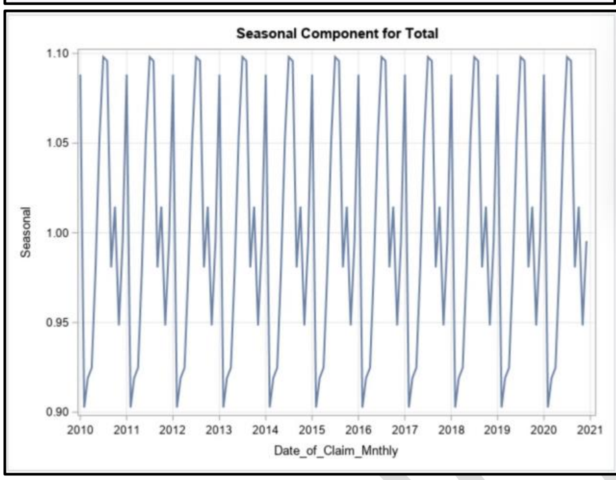

Series Values for Total
Total
Date_of_Claim_Mnthly
400
500
600
700
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021

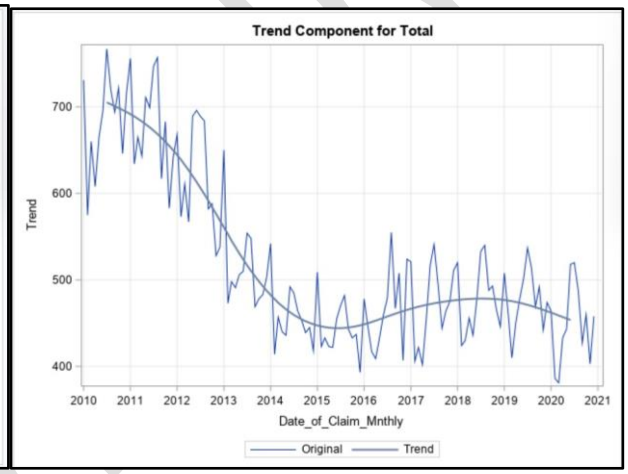

Seasonal Cycles for Total
Total
Season
400
500
600
700
2
4
6
8
10
12

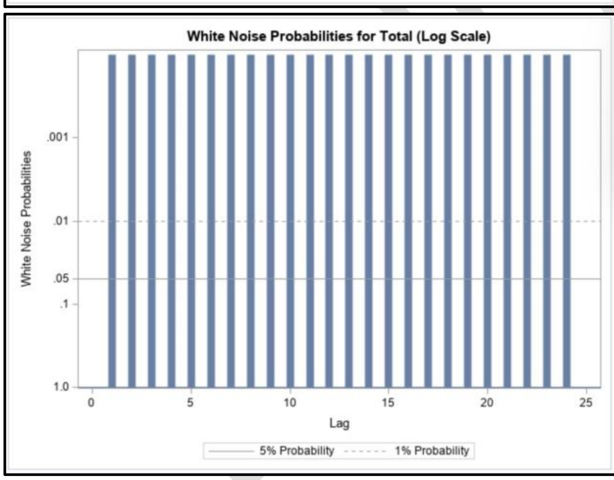

Correlations for Total
ACF
PACF
IACF
White Noise Prob
Lag
Lag
Lag
Lag


Cycle Component for Total
Cycle
Date_of_Claim_Mnthly


Irregular Component for Total
Irregular
Date_of_Claim_Mnthly


Seasonal Component for Total
Seasonal
Date_of_Claim_Mnthly


Trend Component for Total
Trend
Date_of_Claim_Mnthly
Original
Trend

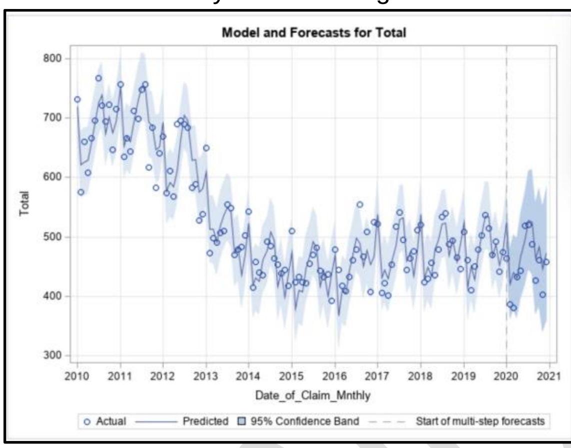

White Noise Probabilities for Total (Log Scale)
White Noise Probabilities
Lag
5% Probability
1% Probability

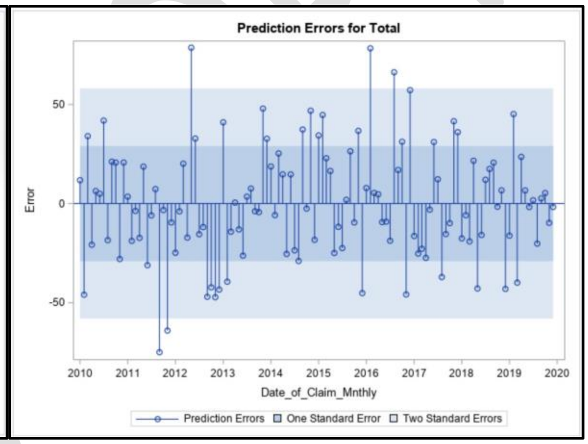
Building the xWeather:

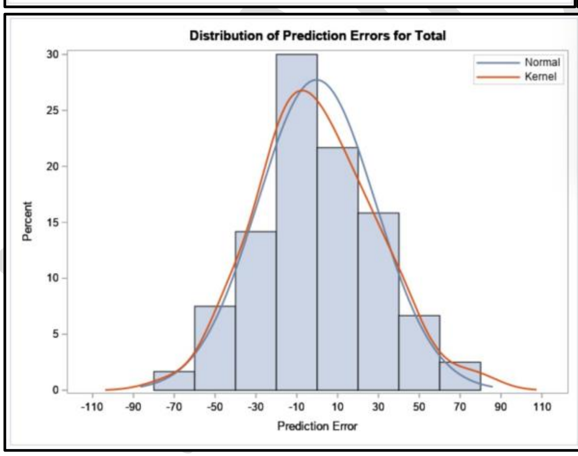
Given there is seasonality and trend in this dataset the first 2 models the team explored are (1) Winters additive method and (2) Winters multiplicative model.

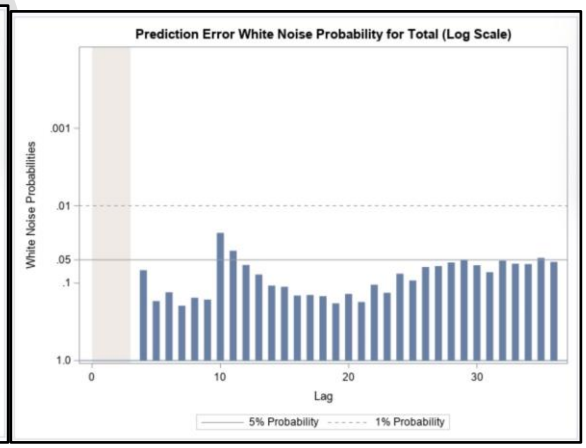
Setting up the ESM Winters additive model, the team started with the forecasting task "Modeling and Forecasting", selecting the project_mnth_ncat_nhlotr_nwthr table as the data. "Total" as the dependent variable, date_of_claim_mnthly as the time element. For the model, selected the ESM and selected Winters additive method and ran all plots. On the options tab, select 12 periods to forecast as our data is monthly and 12 periods for the hold back. For the output, create fit statistics was checked and called it "ESM_Wint_Add".

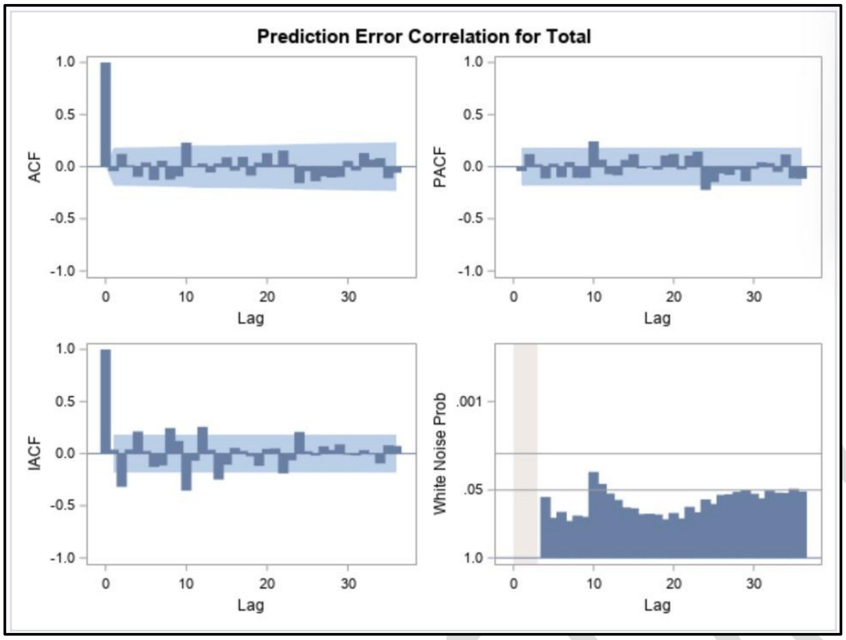
Overall the model did well, with errors that were normally distributed, mostly within 1-2 standard deviations and were white noise. The ACF and PACF on the errors did not show signs of autocorrelation, with only a few lags on the IACF being somewhat significant. Looking at the series and forecasts, the model seems to do reasonably well following the actual values. Plots for this model can be seen below:







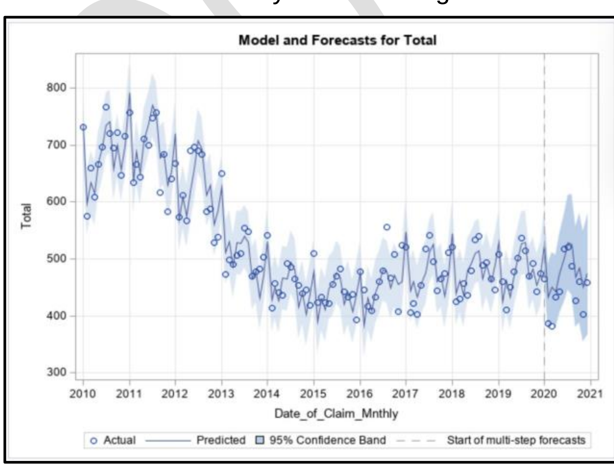


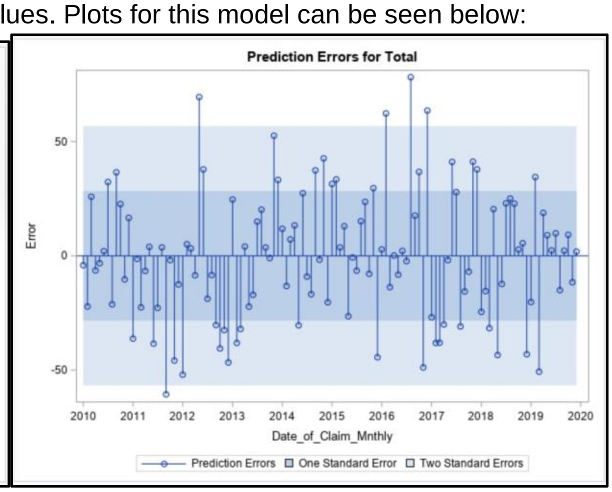


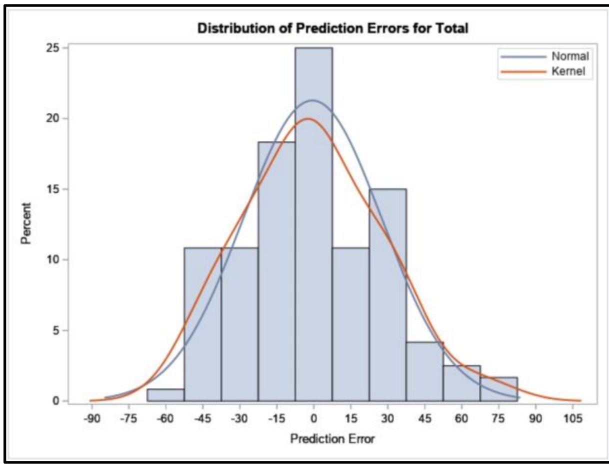
As mentioned, we also created a fit statistics output for this model.
SBC - Fit: 819.661; Forecast: 85.20
AIC - Fit: 811.299; Forecast: 85.20
RMSE - Fit: 28.66; Forecast: 34.809
MAPE - Fit: 4.46; Forecast: 6.958

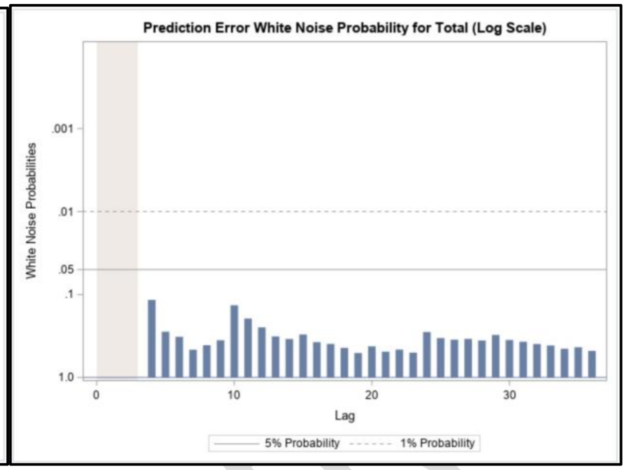
Setting up the Winter's multiplicative model exercise, the team started with the forecasting task "Modeling and Forecasting", selecting the project_mnth_ncat_nhlotr_nwthr table as the data. "Total" as the dependent variable, date_of_claim_mnthly as the time element. For the model, selected the ESM and selected Winter's multiplicative model and ran all plots. On the options tab, select 12 months to forecast as our data is monthly and 12 months for the hold back. For the output, create fit statistics was checked and called it "Esm_wint_mult".

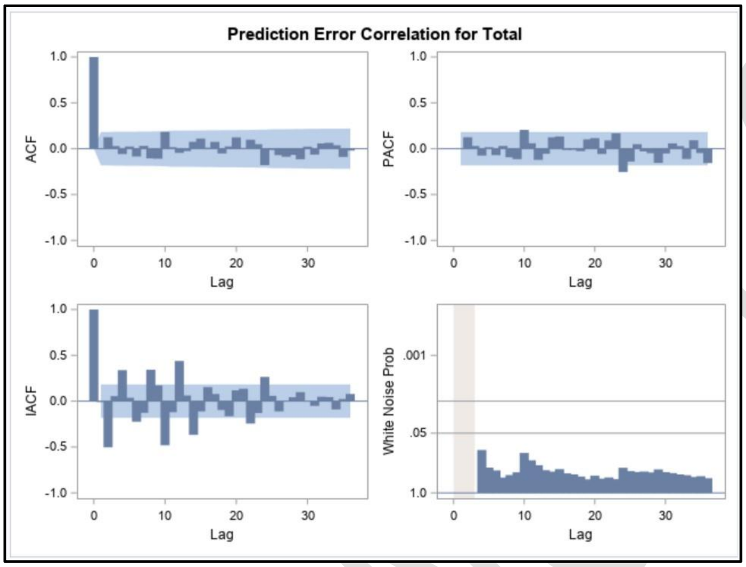
Overall the model did well, with errors that were normally distributed, mostly within 1-2 standard deviations and were white noise. The ACF and PACF on the errors did not show signs of autocorrelation, with only a few lags on the IACF being somewhat significant. Looking at the series and forecasts, the model seems to do reasonably well following the actual values. Plots for this model can be seen below:











As mentioned, we also created a fit statistics output for this model.
SBC - Fit: 814.178; Forecast: 88.04
AIC - Fit: 805.815; Forecast: 88.04
RMSE - Fit: 28.01; Forecast: 39.191;
MAPE - Fit: 4.36; Forecast: 7.964

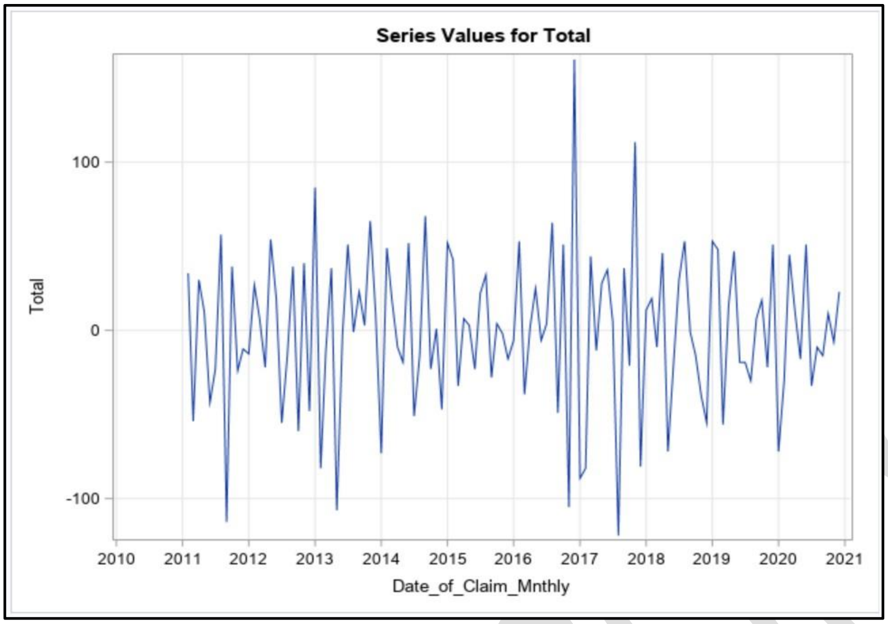
Between these 2 models, the **Winters Multiplicative Model** did a slightly better job and therefore was the model that will be chosen for the xWeather model. The multiplicative model had slightly lower AIC & SBC with slightly worse RMSE & MAPE. We prioritized the AIC & SBC. We also noticed that the multiplicative model did much better in capturing the signal in the data, leading to a better white noise probability plot.

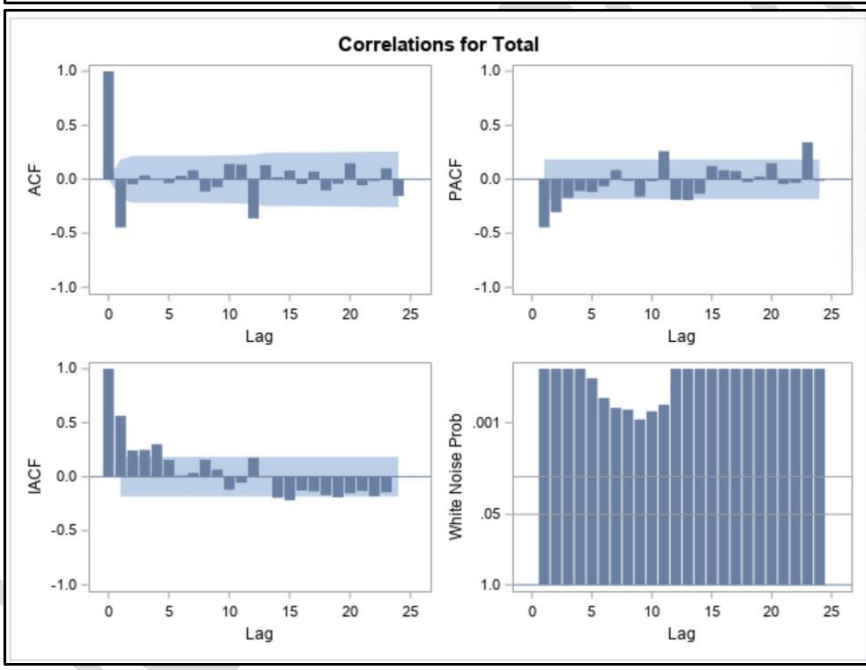
ARIMA models were considered however they did not make the final model selection process because we could not get comfortable the significance levels when selecting the different moving average order level. In addition, the autocorrelation plots looked decent but the white noise plots still show significance at different lags. We tried multiple different combinations of autoregressive and moving average assumptions. In addition, for each combination we tried, the SBC and the AIC remained over 1,000 when compared to the ESM models, which were under 900. Ultimately, we were not going to choose the ARIMA model. Below is an example for the most successful combination. The data had to be transformed with 1 simple differencing and 1 seasonal differencing because 1 seasonal differencing did a good job getting rid of seasonality, but the trend remained. Adding a 1 simple differencing removed the trend.

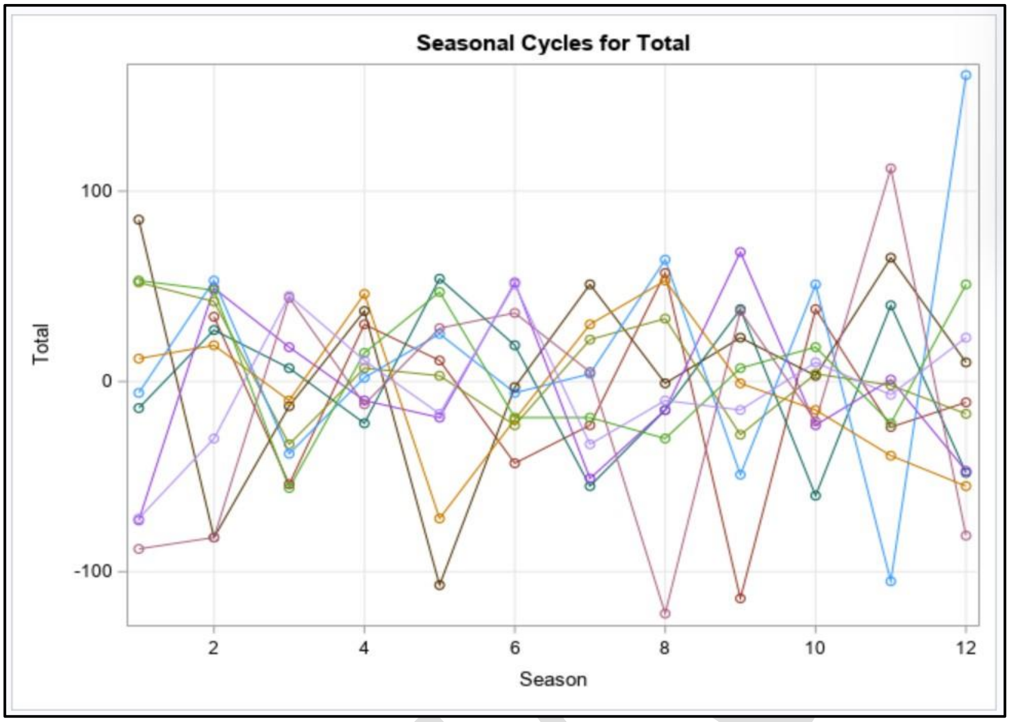
To generate the data needed for an ARIMA model the team used the forecasting task called "Time Series Data Preparation". Within this task we leveraged the table project_mnth_ncat_nhlotr_nwthr. The task was set up to use "Total" (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly, and the transformation of "Total" to accumulation by means of summation, simple differencing = 1 and seasonal differencing = 1.

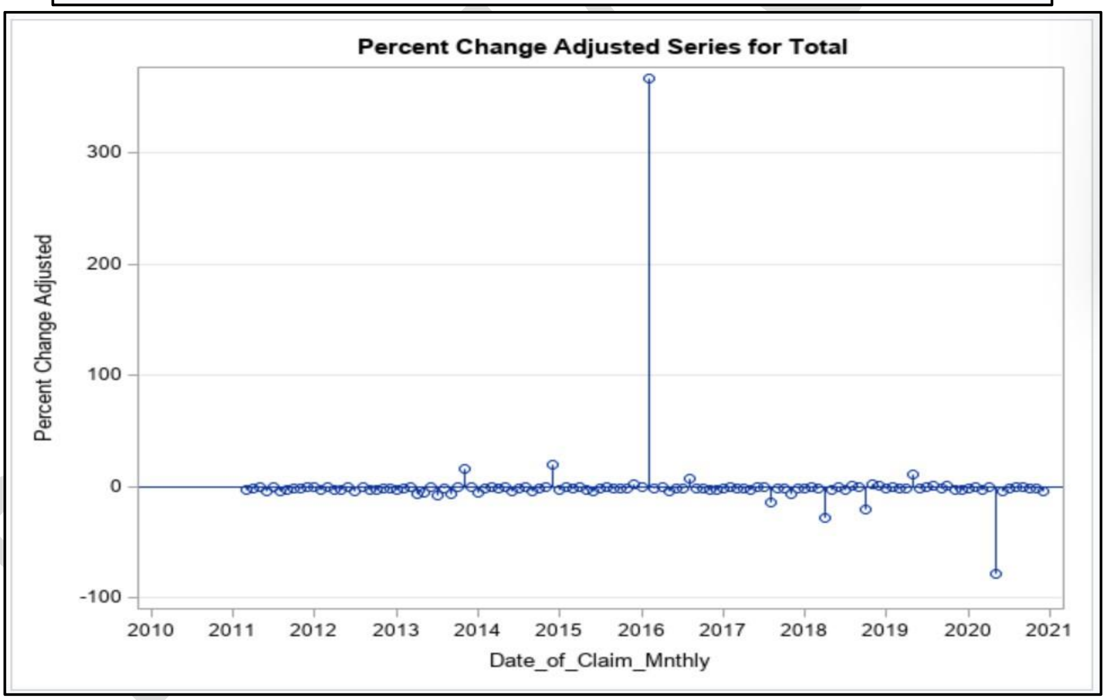
To look at the result of 1 simple differencing & 1 seasonal differencing on the xWeather data we used the forecasting task called "Time Series Exploration ". The task was set up: "Total" (claim volume) as the dependent variable, Date_of_Claim_Mnthly as the Time ID with the interval set to monthly. We ran a gambit of plots and reports.

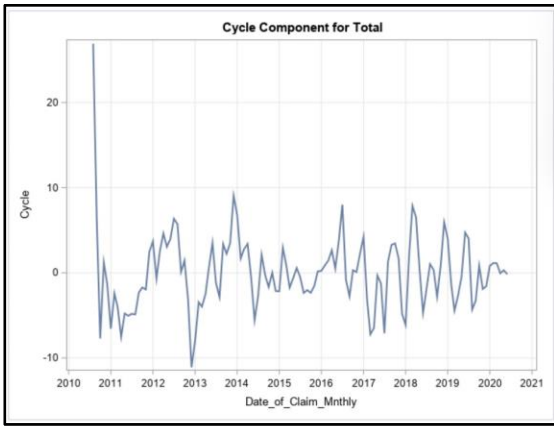
It can be seen that the differencing applied to the data took care of the trend which has essentially been eliminated. It also took care of the seasonality, which has been reduced and nearly eliminated. There still remains no cycle or irregular component. However, the correlation plots do still show autocorrelation and the data is not white noise, meaning a model could still be made to capture that remaining signal. We focused on autoregressive <=2 because the IACF and the PACF had significant spikes in lags 1 & 2 and we looked at moving average <=1 because the ACF has a significant spike in lag 1. See plots below.

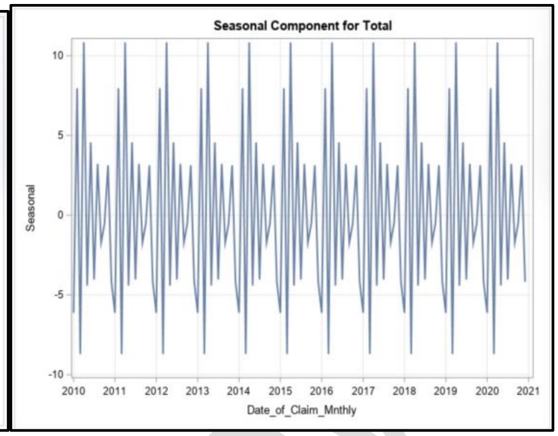

Series Values for Total
Total
100
0
-100
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021
Date_of_Claim_Mnthly

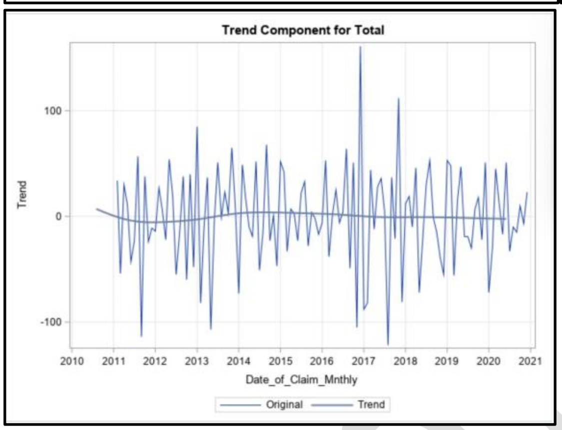

Correlations for Total
ACF
PACF
IACF
White Noise Prob
1.0 0.5 0.0 -0.5 -1.0
.001 .05 1.0
0 5 10 15 20 25
Lag


Seasonal Cycles for Total
Total
Season
100
0
-100
2
4
6
8
10
12


Percent Change Adjusted Series for Total
Percent Change Adjusted
Date_of_Claim_Mnthly
300
200
100
0
-100
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021







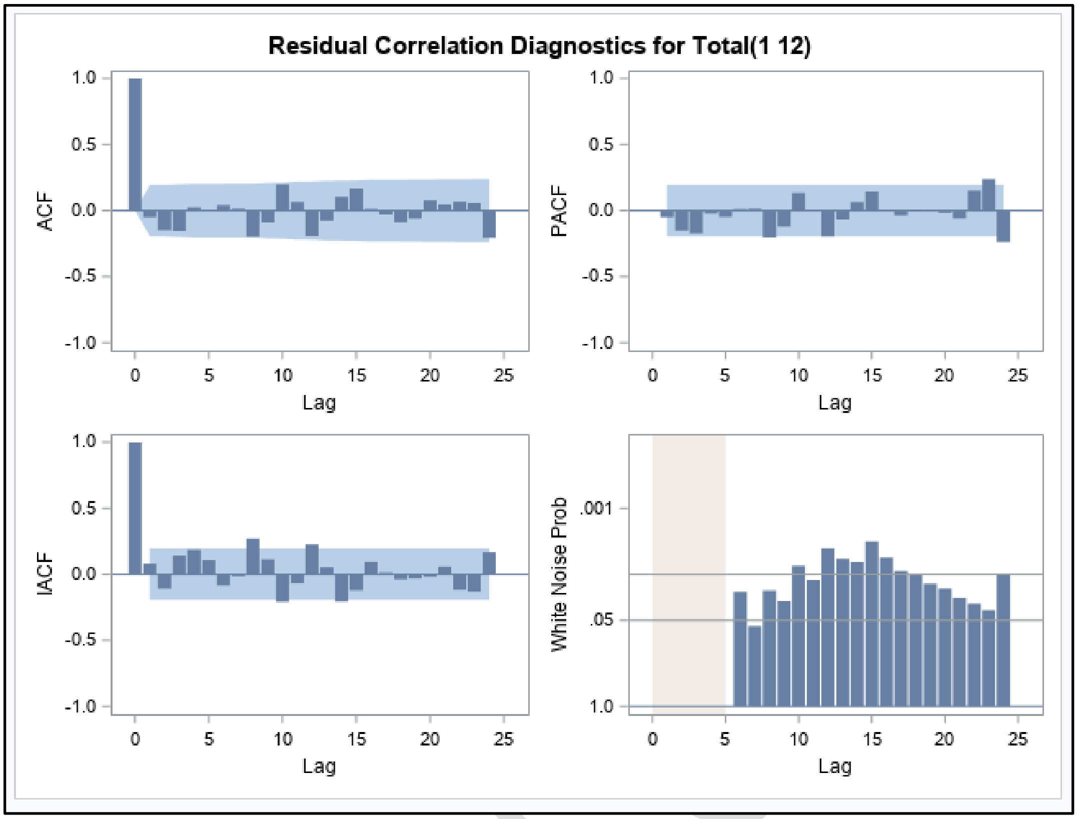
Setting up the best ARIMA model, the team started with the forecasting task "Modeling and Forecasting", selecting the PROJECT_MNTH_NCAT_NHLOTR_NWTHR_D table as the data. "Total" as the dependent variable, date_of_claim_mnthly as the time element. For the model, selected the ARIMA with and selected p=2, d=1, q=1, P=2, D=1, Q=0 and ran all plots. We experimented with seasonal autoregressive picks because of the spikes in the PACF happening every 12 lags. On the options tab, select 12 months to forecast as our data is monthly and 12 months for the hold back.

Originally the thought was this was going to be a moving average selection because the PACF and the IACF are decaying from lag 0 but through our trial and error the result showed that the model we built disliking the moving average selections because it kept failing the p-test were significant. Below is the Maximum likelihood estimation for our best effort:

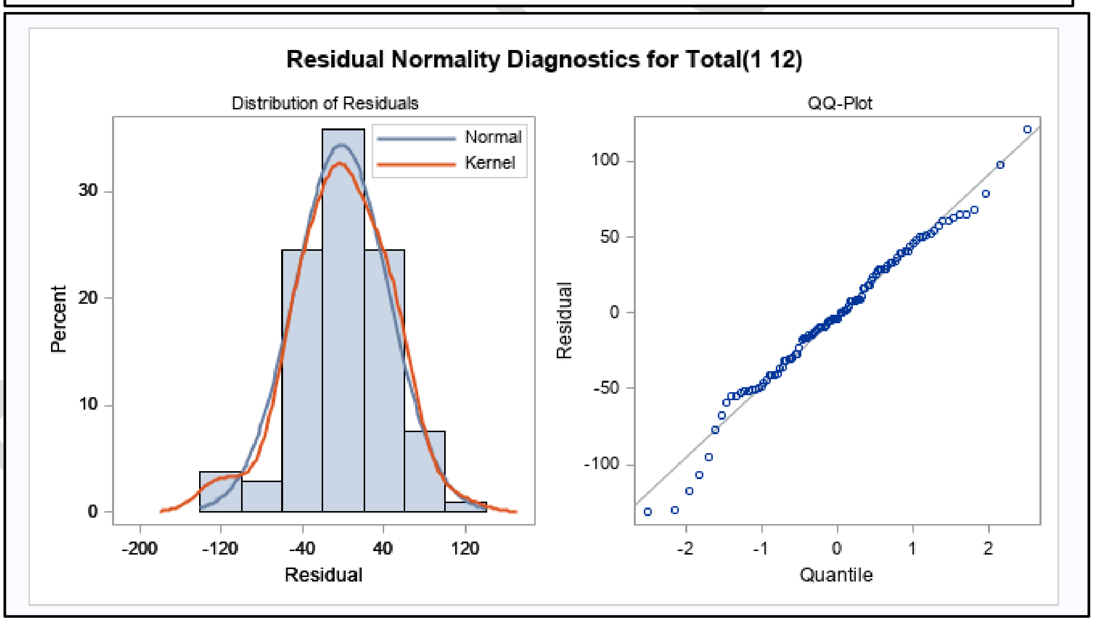
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.04324	0.07797	-0.55	0.5792	0
MA1,1	0.99987	19.89903	0.05	0.9599	1
AR1,1	-0.61926	0.09147	-6.77	<.0001	1
AR1,2	-0.26607	0.09435	-2.82	0.0048	2
AR2,1	-0.86325	0.08055	-10.72	<.0001	12
AR2,2	-0.51352	0.08268	-6.21	<.0001	24

Constant Estimate	-0.19374
Variance Estimate	2265.098
Std Error Estimate	47.59305
AIC	1145.465
SBC	1161.446
Number of Residuals	106

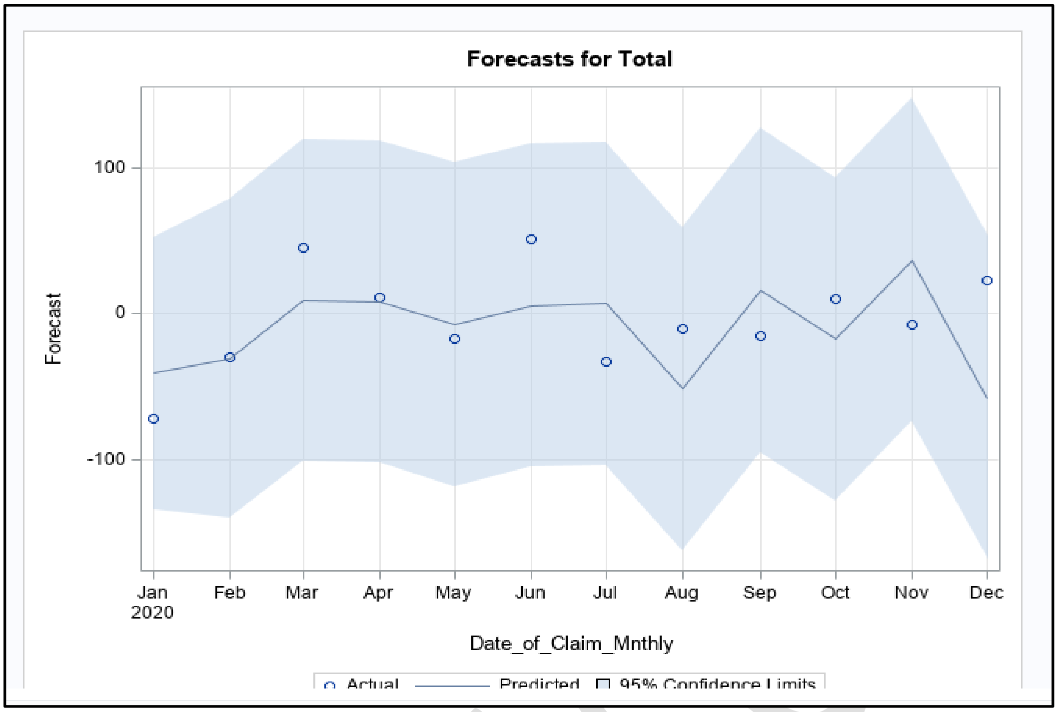
You can see that the model did not like MA(1,1) with a 0.9599 for the |t| test, however if we removed the MA selection the autocorrelation in the residuals came back and the white noise test looked worse with more significant bars. Below are the results of the best options we picked with the lowest SBC and AIC of 1161.446 and 1145.465, respectively. The residuals are normally distributed, the ACF/PACF/IACF for the residuals look reasonable and not significantly but the white noise is elevated. This would have been a contenting model if we felt better about the t-test for MA but since it was outside our comfort level and more importantly the SBC could not match the ESM models, we did not pursue future refinement. Below the plots, we provided an example of removing the MA selection.



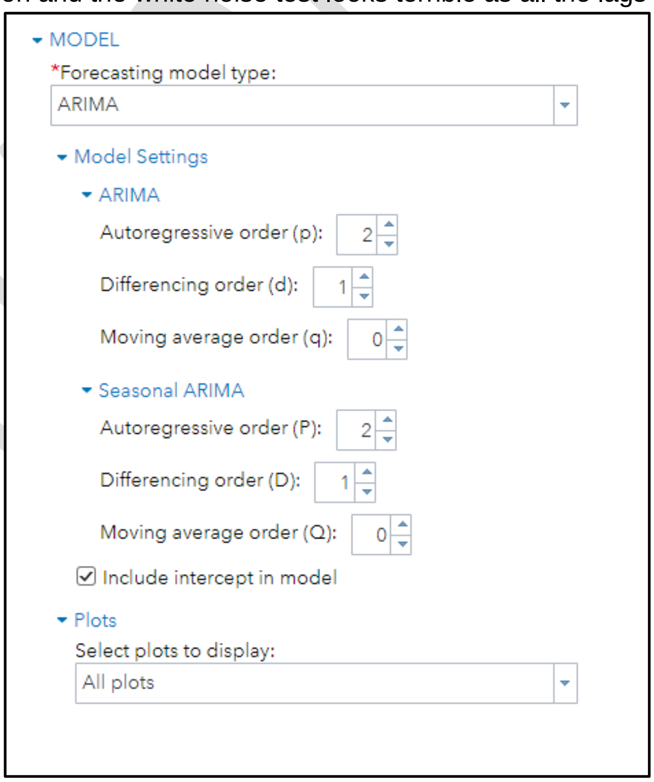
Residual Correlation Diagnostics for Total(1 12)
ACF
PACF
IACF
White Noise Prob
Lag



Residual Normality Diagnostics for Total(1 12)
Distribution of Residuals
Normal
Kernel
Percent
Residual
QQ-Plot
Quantile

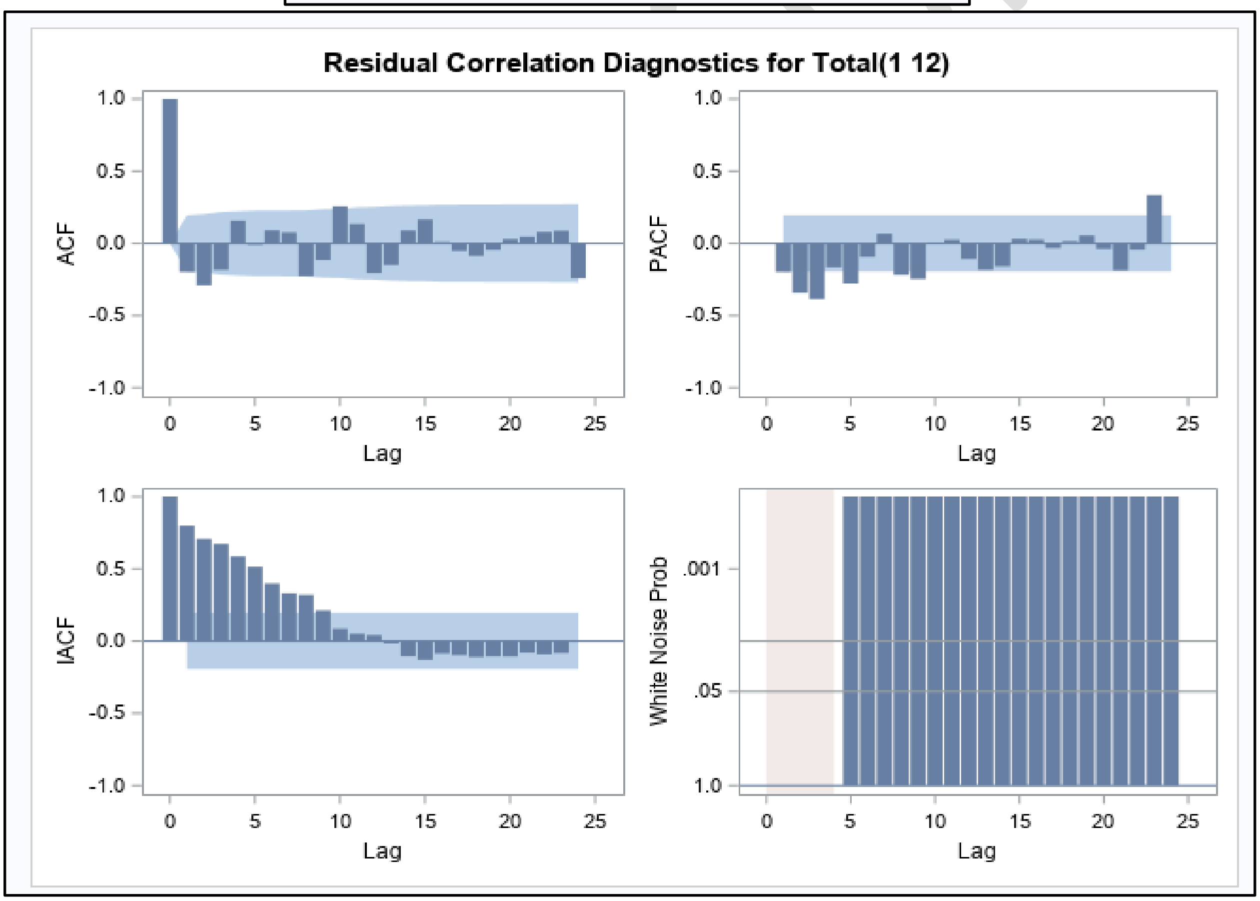


Example of removing the MA selection from our best attempt. The SBC and AIC deteriorate but the AR selections both p and P show they are not significant and pass the t-test. However, the PACF and IACF are showing autocorrelation and the white noise test looks terrible as all the lags are significant.



Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.08146	1.04673	0.08	0.9380	0
AR1,1	-0.98651	0.08050	-12.25	<.0001	1
AR1,2	-0.49912	0.08356	-5.97	<.0001	2
AR2,1	-0.92152	0.08277	-11.13	<.0001	12
AR2,2	-0.51660	0.08543	-6.05	<.0001	24

Constant Estimate	0.493662
Variance Estimate	3915.361
Std Error Estimate	62.57284
AIC	1196.772
SBC	1210.089
Number of Residuals	106



CONCLUSION:

The four next step actions and recommendation for ABC insurance:

1. After reviewing the model results our next steps are to explore the seasonality within the non-weather data set with loss analytics to have a better understanding of the driving forces and to validate frozen pipes are not being miscoded given the January volumes.
2. Work with Research and Development (R&D) and reserving actuaries to peer review the models.
3. Recommend CAT (non-hail) forecasts be handled using industry standard models like RMS (https://www.rms.com) or AIR (https://www.air-worldwide.com/) for hurricane, earthquake, terrorism, and wildfire perils.
4. Recommend a simple 5-year average to handle forecasting "All Other" activity**.**

REFERENCES:

1. George Fernandez, Marc Huber, Jay Laramore, Danny Modlin, and Chip Wells, ISBN 978-1-64295-144-8, Time Series Modeling Essentials Course Notes
2. https://www.natlawreview.com/article/pandemic-vs-policyholder-covid-19-and-business-interruption-coverage-claims