Boundary Box Failure Analysis

Anglian Water Services Limited

Project number: 60702454

28 March 2023

Delivering a better world

Quality information

Prepared by	Checked by	Verified by	Approved by
Cheuk Lun Bernie Tse Graduate, Asset Management	Simon Gee Technical Director, Geospatial & Asset Management	Alex Mortlock Head of Asset Management & Advisory	Alex Mortlock Head of Asset Management & Advisory

Revision History

Revision	Revision date	Details	Authorized	Name	Position
0	28/03/2023	Draft for client approval			
Final	18/04/2023	Approved by client		Simon Gee	Technical Director

Distribution List

# Hard Copies	PDF Required	Association / Company Name

Prepared for:

Tom Swain
PR24 Portfolio Manager - Water
Anglian Water Services Limited
Thorpe Wood House
Thorpe Wood
Peterborough
PE3 6WT

Prepared by:

Cheuk Lun Bernie Tse
Graduate, Asset Management
AECOM Limited
Winslade House
Winslade Park
Manor Drive
Clyst St Mary
Exeter, EX5 1FY
United Kingdom

T: +44 1392 663200
aecom.com

Table of Contents

1. Introduction

Anglian Water Services (AWS) have identified a potential emerging issue related to the increasing failure rate of boundary boxes and stop taps. Before pursuing a potential cost factor adjustment claim with Ofwat, AWS are seeking independent analysis of the job data to inform their approach.

AECOM were commissioned to undertake analysis of boundary box job data to inform AWS' understanding of the apparent acceleration in stop tap and boundary box failures.

The data available was a simple count of the number of reactive jobs undertaken on boundary boxes and stop taps, by month. AWS also provided a spreadsheet with their analysis of the data, including a forecast using a time series function provided in Excel.

A number of commonly used time series forecasting methods were reviewed and used to analyse the data and the results compared to the forecast generated using the Excel function.

2. Data Processing

A simple count of the number of reactive jobs undertaken by month was used for this analysis. No further data was available regarding the types of jobs undertaken or the types or ages of the assets involved.

A forecast produced by AWS using the Excel inbuilt function FORECAST.ETS was provided as shown in Figure 1 below.

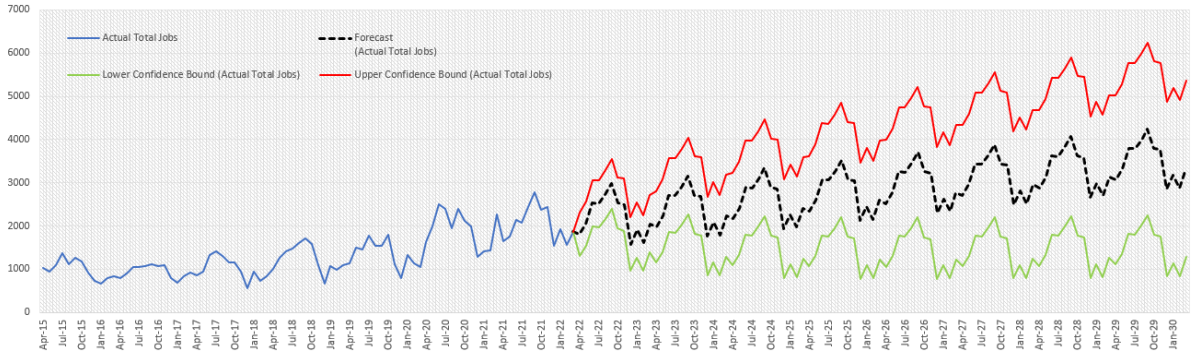


Figure 1. AWS Excel Prediction

The analysis was undertaken using the R scripting language with library 'fpp2' and 'forecast'. In total 5 time series forecasting models were tested in R.

Name	Description
ETS AAA	Exponential Smoothing (additive error, additive trend and additive seasonality)
ETS MAM	Exponential Smoothing (multiplicative Holt-Winters' method with multiplicative errors)
ARIMA (0,1,1)(0,1,1) CSS	Auto Regressive Integrated Moving Average (conditional sum of squares)
ARIMA (0,1,1)(0,1,1) ML	Auto Regressive Integrated Moving Average (Maximum-likelihood estimation)
Holt-Winters	Holt's method with the ability to capture seasonality

Table 1. Description of time series forecasting models

3. Analysis

Initial analysis of the data was performed using the 'decompose' function. Decomposing a seasonal time series involves separating the time series into three components of seasonal, random and underlying trend.

There is a clear upward trend in the number of reactive jobs and a strong seasonal pattern as shown in Figure 2. Therefore only time series forecast models suitable for seasonal data were considered.

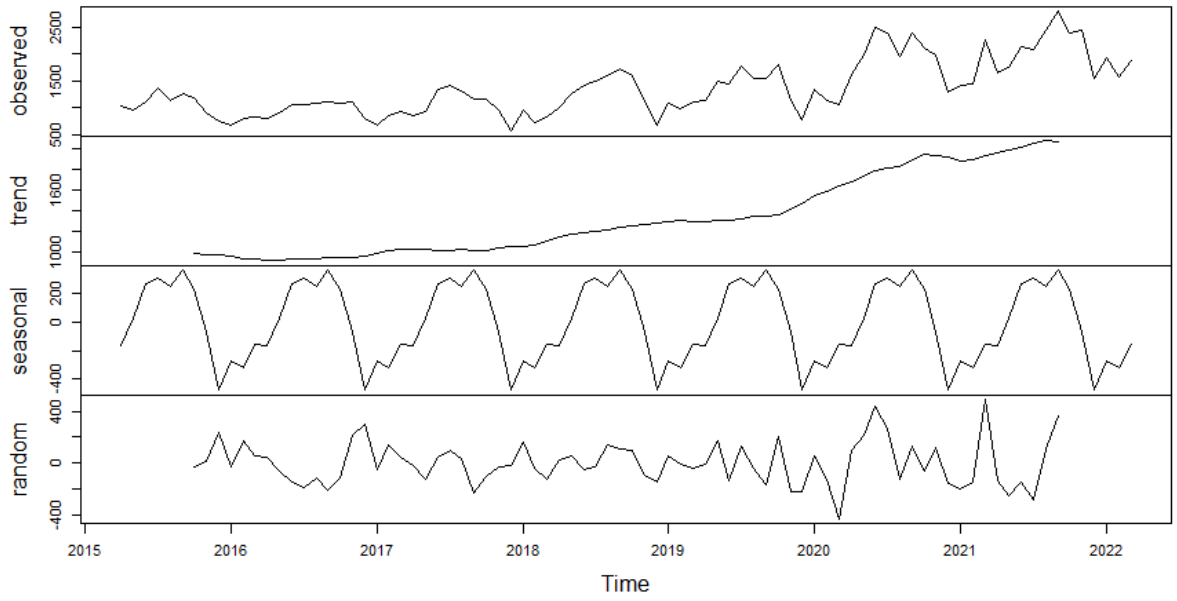


Figure 2. Decomposition of the raw data

In order to compare the relative accuracy of the different models, the Mean Absolute Percentage Error MAPE measure was used. MAPE is one of the most comonly used error metrics to compare the forecast performance between different time series models.

The MAPE is an indicator of Demand Forecasting performance that measures the size of the (absolute) error in percentage terms. The fact that a magnitude of the percentage error is estimated makes it an indicator frequently used by forecasters due to its easy interpretation. MAPE is a relative error measure that uses absolute values to keep the positive and negative errors from canceling one another out and uses relative errors to enable comparison of forecast accuracy between time-series models.

For the initial analysis, the total number of jobs from the end of the observation period to the end of AMP8 (2022-2030) was predicted.

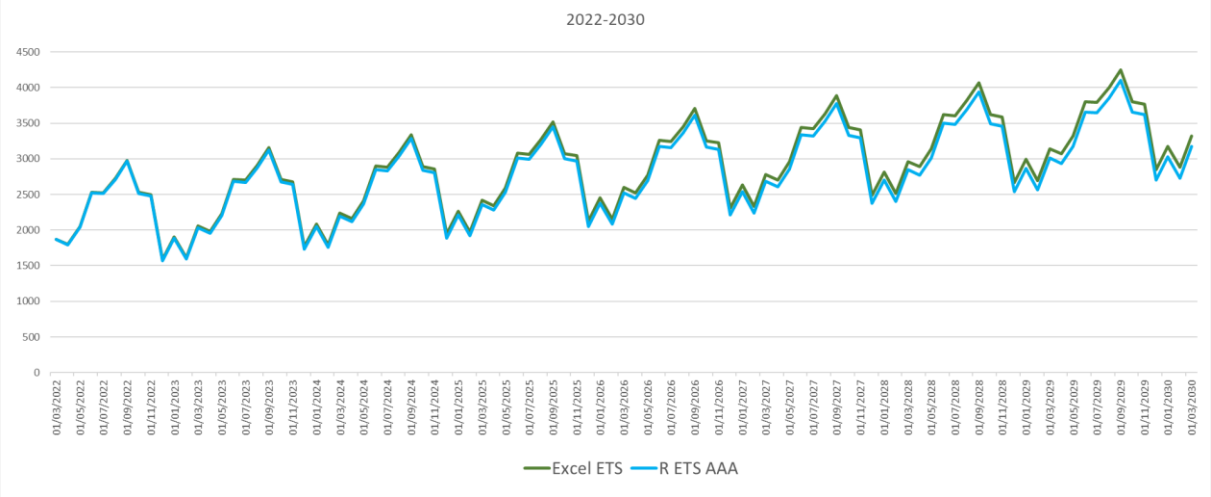
3.1. ETS AAA (Excel parameters)

The Exponential Time Smoothing (ETS) Additive error, Additive trend, and Additive seasonality (AAA) model used by Excel was reproduced in R using the "forecast" package. There is limited documentation about this inbuilt function in Excel. The alpha, beta and gamma parameter values were extracted from the Excel model and used to reproduce it in R. The predictions produced in R are almost identical to the results from Excel, in both the total number of jobs and the seasonal pattern, which is to be expected. This model resulted in a MAPE of 13.19, meaning that it is 86.8% accurate.

Statistic	Value
Alpha	0.250
Beta	0.001
Gamma	0.749

Table 2. Excel ETS AAA parameters

Figure 3 shows how close the match is between the Excel prediction (green) and that derived in R (blue). Reproducing this forecast was considered an important test of the overall approach to using time-series forecasting models in R.



	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	1.705068	250.0583	185.3695	-2.653336	13.19439	0.709096	0.1806917

Figure 3. ETS AAA in Excel vs. R, using Excel parameters

3.2. ETS AAA(Auto parameters)

Another prediction was also produced using the same ETS AAA model, but with the alpha, beta and gamma values chosen by R. There is a 12% decrease in the total number of jobs predicted compared to the Excel one with 242,403 jobs predicted from 2022 to 2030. But the MAPE for this model is improved to 11.71, suggesting that the Excel parameters in the original model were not the optimal ones for the analysis, as the auto generated parameters give a lower level of error.

```
Smoothing parameters:
  alpha = 0.3345
  beta  = 1e-04
  gamma = 1e-04
```

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	12.9257	198.1867	154.4619	-0.5028771	11.71092	0.5908648	0.07827162

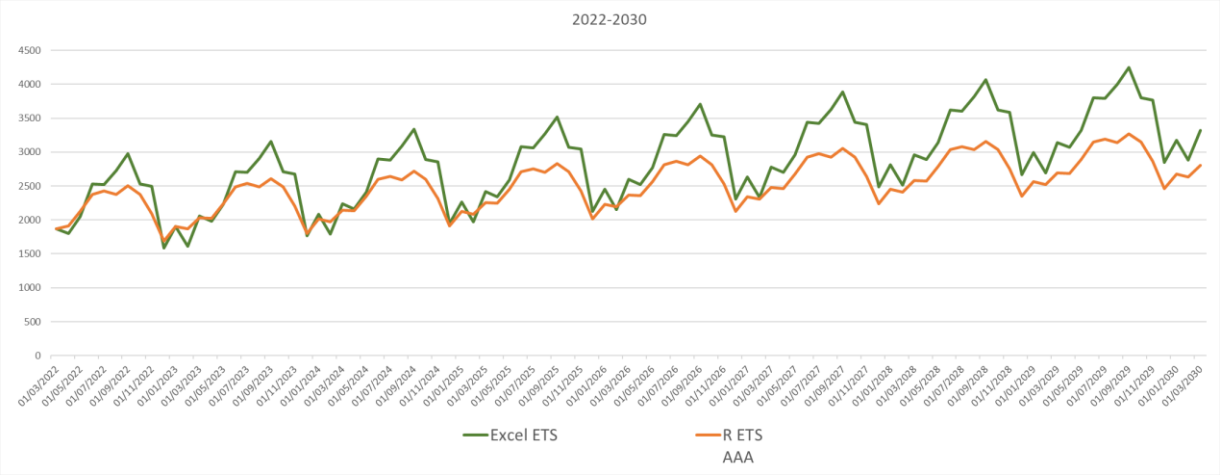


Figure 4. ETS AAA in R, using parameters chosen by R

3.3. ETS MAM

The ETS MAM model represents Mutiplicative error (error as the ratio between the measurement and the truth), Additive trend (a generally ;inear underlying trend) and Multiplicative seasonailty (essentially, increasing widths or heights of seasonal periods over time). In the historical data in Figure 2, there is a clear trend and increasing seasonality. This method is recognised as best suited to data with an underlying trend and with seasonality that increases over time, as it results in a curved forecast that also reproduces the seasonal changes in the data.

This model is chosen by R out of all ETS models due to having the lowest MAPE (10.98). This model returned the highest number of total jobs from 2022 to 2030 predicted at 335,933, a 22% increase compared to the predictions produced in Excel.

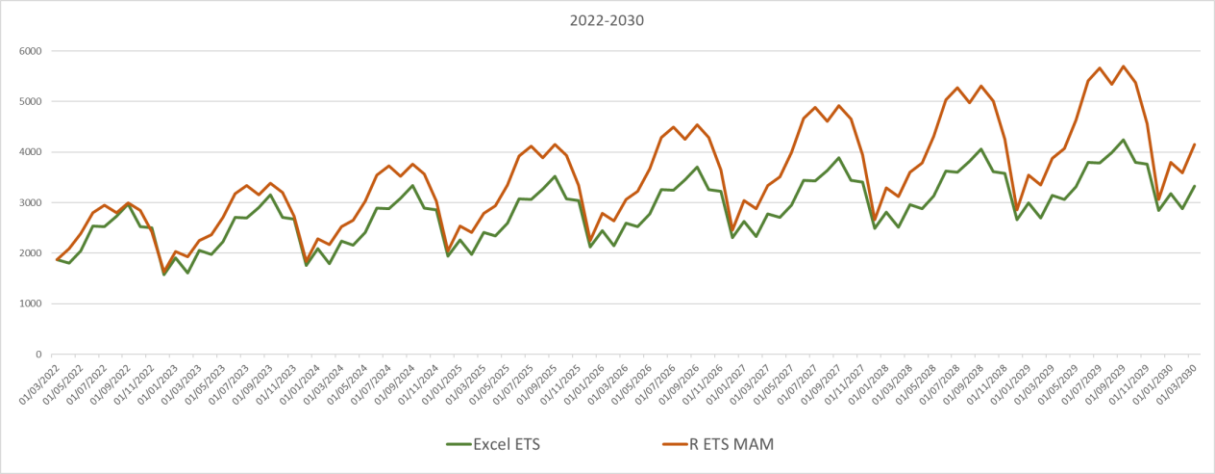


Figure 5. Excel ETS AAA and ETS MAM in R

3.4. ARIMA (Autoregressive Integrated Moving Average Model)

There is a seasonal and a non-seasonal version of the ARIMA model. ARIMA uses autocorrelation, which relates the same set of observation but across different timing (e.g. between rainfall in the summer versus that in the winter). The seasonal ARIMA model (SARIMA) was used due to the obvious seasonality in the observed data. There are multiple parameters in a SARIMA model,

ARIMA(p,d,q)(P,D,Q)m

- p — the number of autoregressive terms
- d — degree of differencing (this manages the trend and seasonal components)
- q — the number of moving average terms
- m — refers to the number of periods in each season
- (P,D,Q) — represents the (p,d,q) for the seasonal part of the time series

These parameters are chosen by R based on the lowest MAPE that can be achieved using the “auto.arima” function. For this dataset the ARIMA (0,1,1)(0,1,1) model was chosen by R. There are multiple fitting methods within the ARIMA model. Both the CSS and ML methods were tested. CSS finds a least squares solution conditional on the assumption that the first, say p observations are exact. ML refers to Maximum Likelihood estimation, a method of estimating the parameters of an assumed probability distribution, given some observed data. This is achieved by maximizing a likelihood function so that, under the assumed statistical model, the observed data is most probable. Similar total numbers of jobs were predicted, and both are close to what Excel predicted with a 2-3% difference.

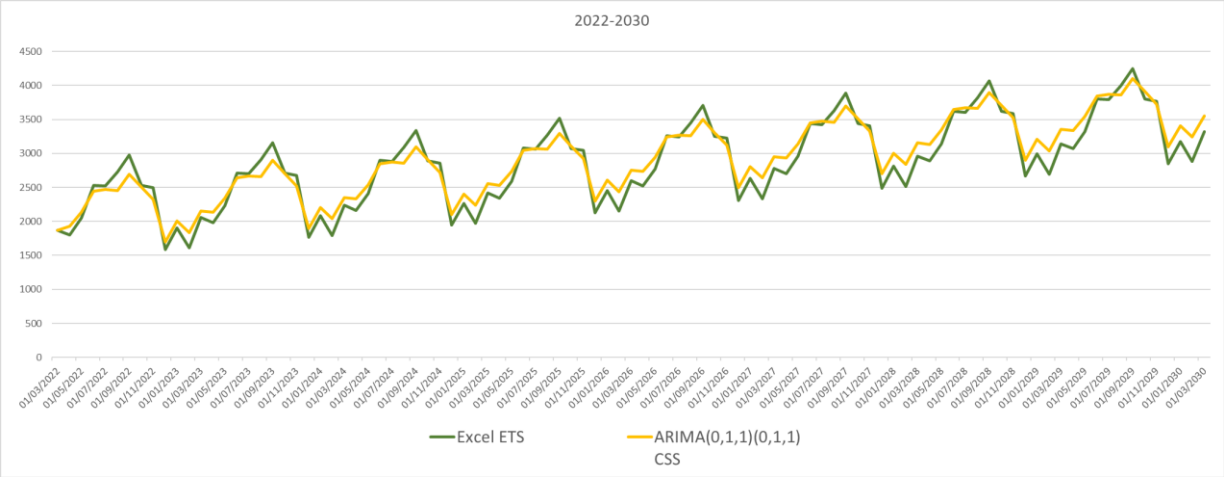


Figure 6. Excel ETS AAA vs. ARIMA,CSS in R

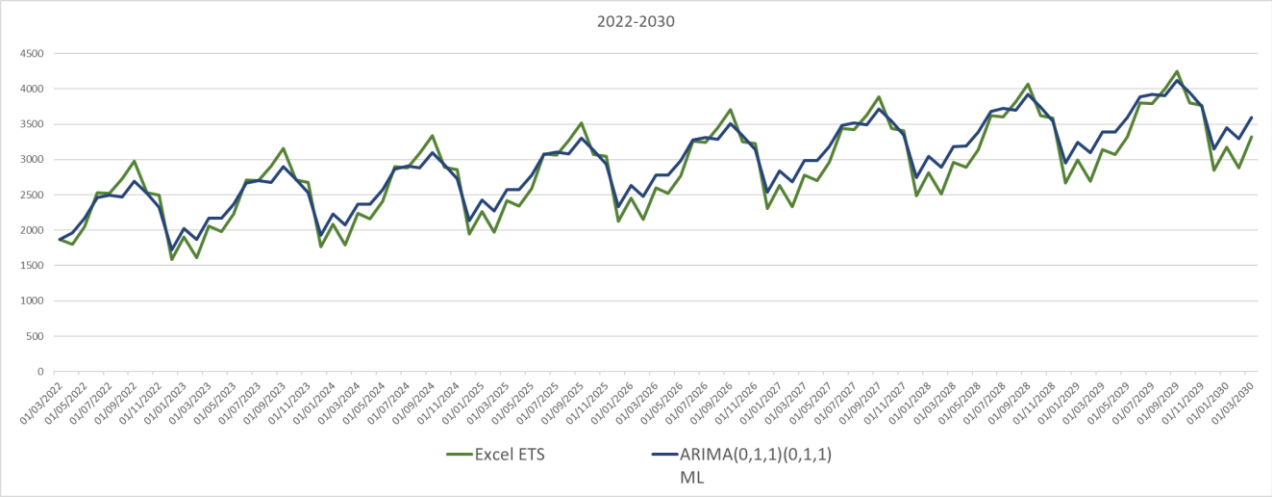
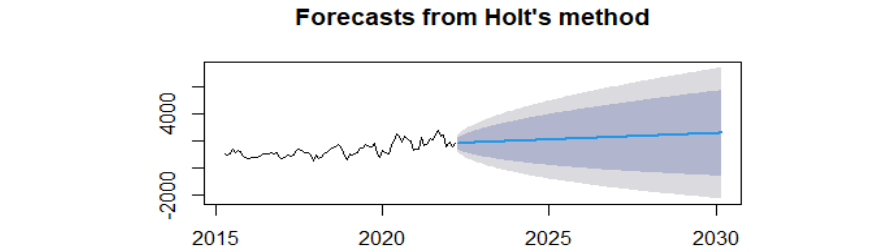


Figure 7. Excel ETS AAA vs. ARIMA,ML in R

3.5. Holt-Winters

The Holt's Method is not a seasonal model as shown in Figure 8 below. Simple exponential smoothing is not appropriate when there is a trend in the data.



	ME	RMSE	MAE	MPE	MAPE
Training set	-0.9359389	299.1821	224.622	-3.295112	17.87738

Figure 8. Holt's Method in R

The Holt and Winters model extends the Holt's method to capture seasonaility. Triple exponential smoothing applies exponential smoothing three times, which is commonly used when there are high frequency signals to be removed from a time series, such as multiplicative seasonality.

This approach has the lowest accuracy out of all methods with a MAPE of 13.55 or 86.4% accuracy. The total number of jobs forecast from 2022-2030 by this model is 317,917, a 16% increase compared to the Excel prediction.

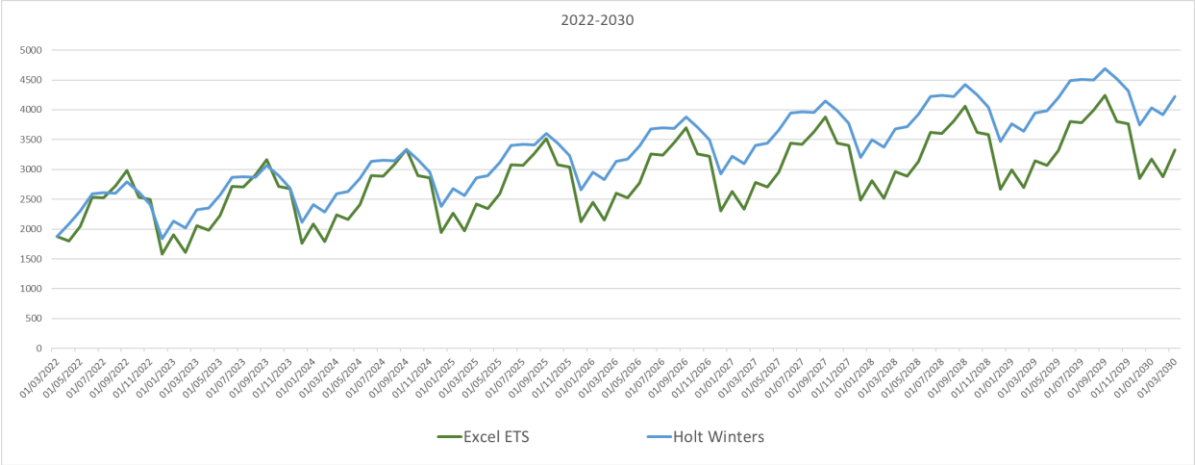


Figure 9. ETS AAA in Excel and Holt-Winters in R

All of the predictions are plotted on the same graph in Figure 10. The ETS MAM method not only has the highest total number of jobs in the period between 2022 and 2030, but also has the most exagerrated seasonailty with the highest peak number of jobs.

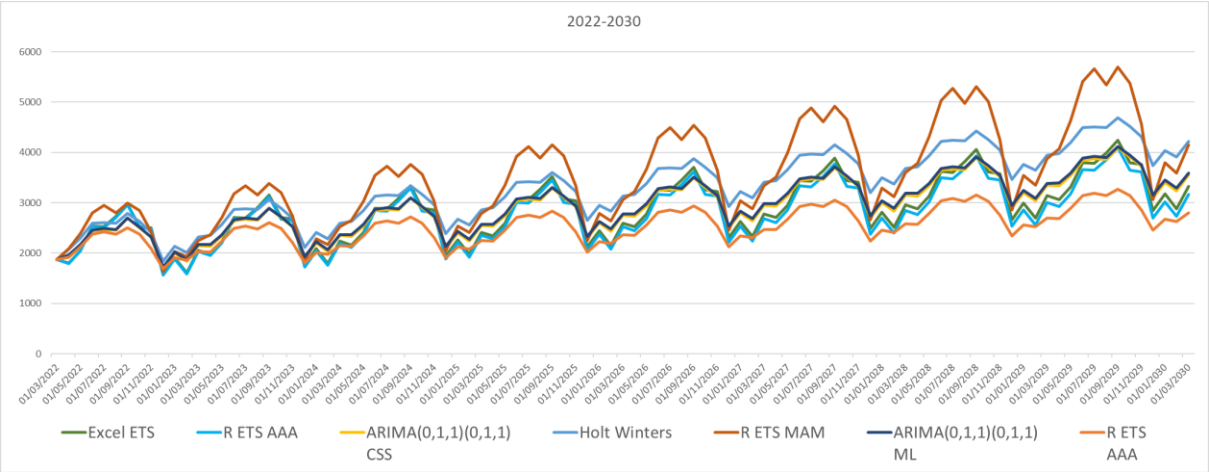


Figure 10. All models comparison

Table three summarises the forecasting models tested, their accuracy using the MAPE measure and the difference between the total number of jobs predicted from 2022 to 2030 in the original Excel version.

Model	MAPE	Difference in Job Count 2022 to 2030
Excel		
R ETS AAA (Excel ABG)	13.19	-3%
ARIMA R CSS	11.69	2%
Holt-Winters	13.55	16%
R ETS MAM	10.98	22%
ARIMA R ML	11.08	3%
R ETS AAA	11.71	-12%

Table 3. MAPE of all models and difference from Excel predictions

All models follows a similar seasonal pattern where the total number of jobs during winter months are lower than in the summer months. A plot of the trendline gives a better represenation of the total number of jobs predicted in the time period.

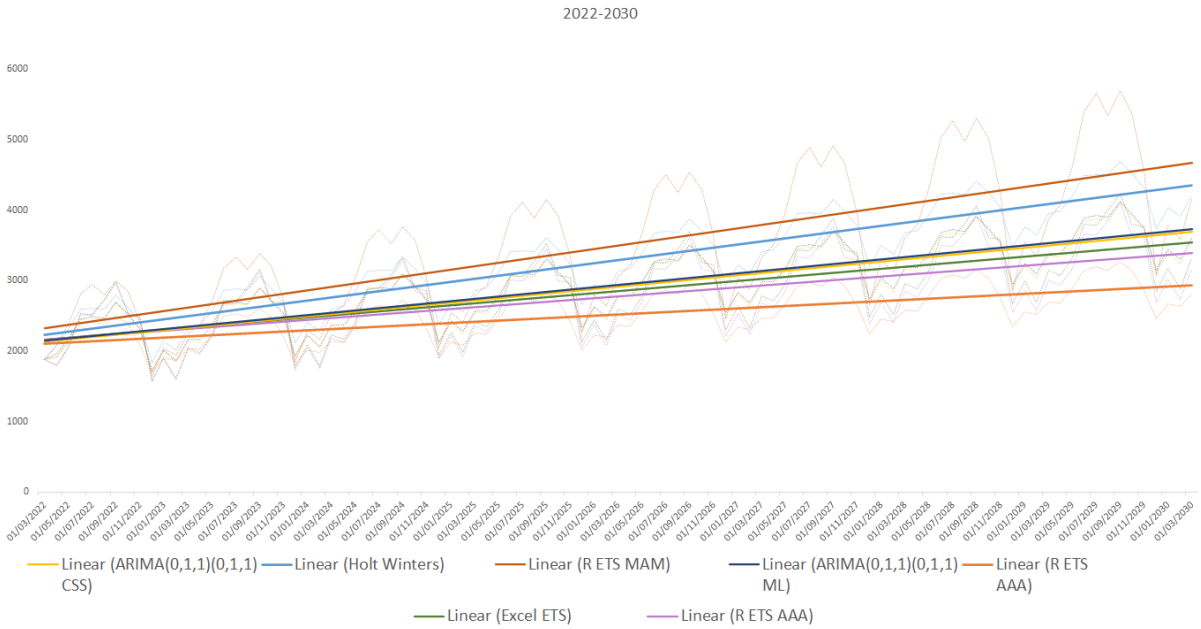


Figure 11. All models trendline plot

3.6. Extended forecast to 2040

The predictions were extended to the end of AMP10 (2040). This does not affect the statistical accuracy of the different models, but the differences in the total number of jobs predicted are exaggerated.

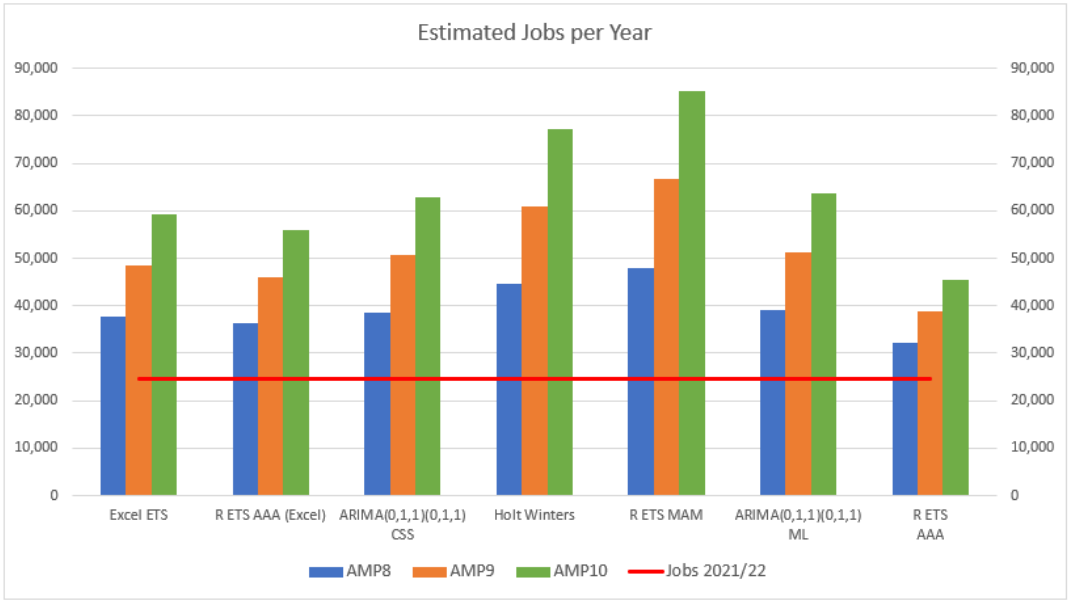


Figure 12. Predictions extended to end of AMP 10

3.7. Effect of prior history

The data for the number of jobs before April 2015 is not available. In order to explore the potential impact of this prior job history, the dataset was extended using an assumed flat rate of 1,000 jobs per month for the ten years. Overall, the total number of jobs predicted was similar to those made without this assumed flat history, across all model options.

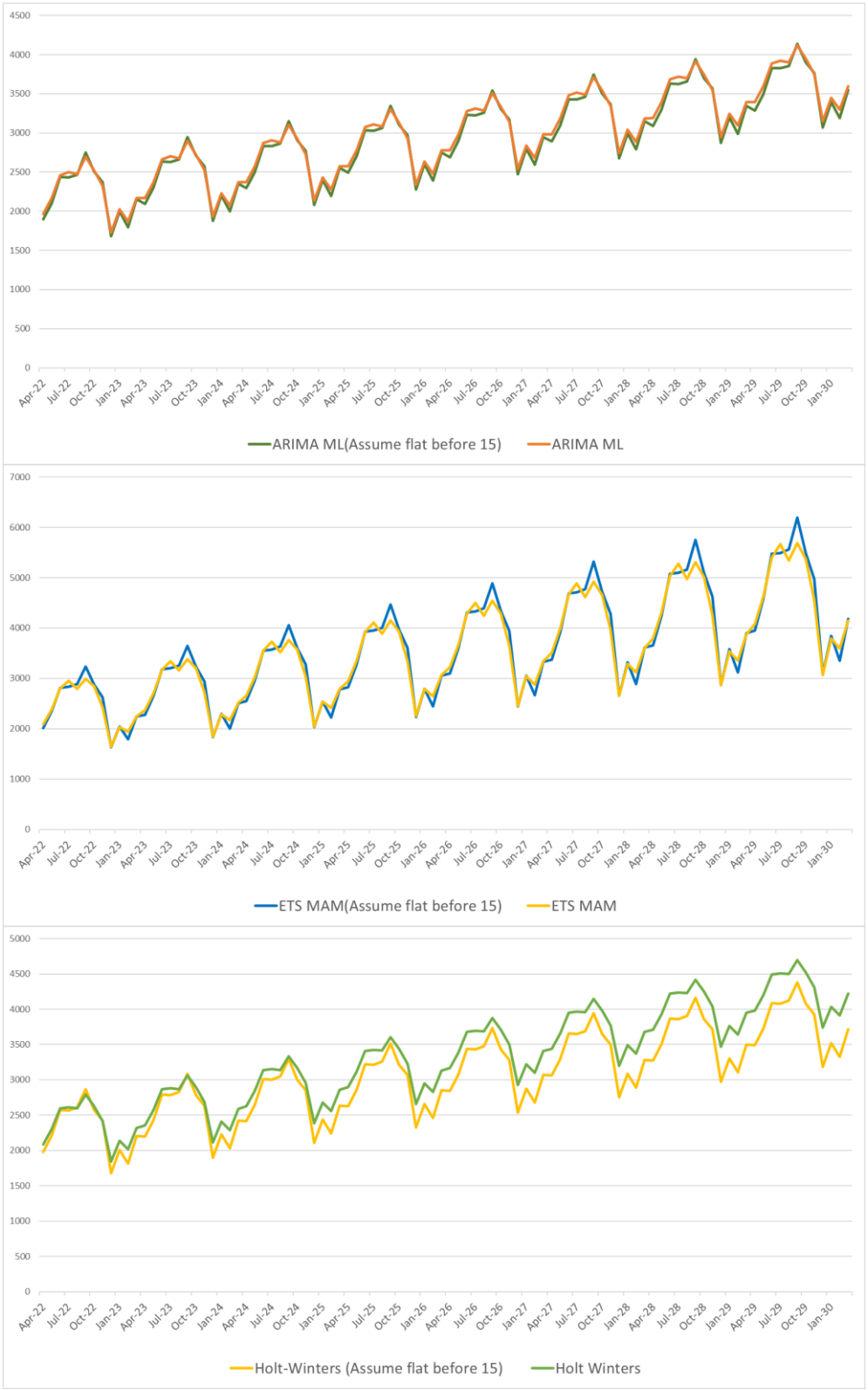


Figure 13. Plots of predictions with extended prior jobs

3.8. Additional Verification

In order to provide additional verifications to the forecast models, a “train-test-split” method was used. The training set contains data from April 2015-March 2019. The test set contains data from April 2019 to March 2022. The predicted number of jobs from the different models were compared against the test set. In both the ARIMA and ETS MAM models, the total number of jobs predicted are less than the actual total number of jobs, with differences ranging from -6 to -11%. This is caused by the steeper increase in trend as shown in Figure 2 at the start of the report. The ideal testing time period is around 5-8 years if the predictions required are from 2022 to 2030. Due to limited time period of the data provided, this is not possible.

	ARIMA	ETS MAM
Total Number of Jobs Predicted (19-22)	60,260	56,878
Difference to Historical Data	-6%	-11%

Table 4. Train-test-split results (Train 4yrs, Test 3yrs)

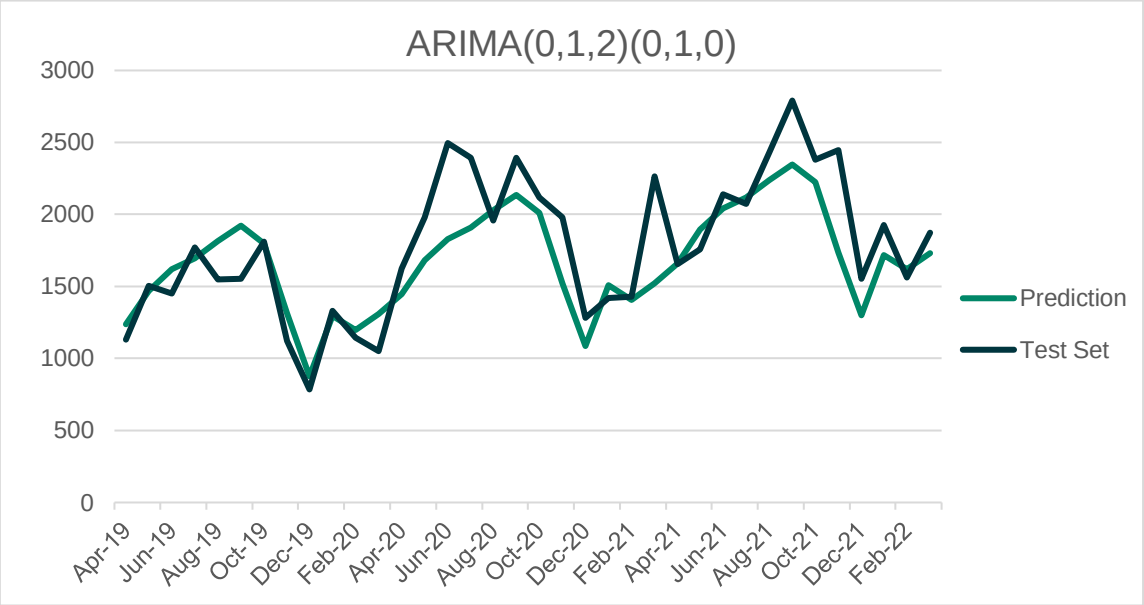


Figure 14. Verification of ARIMA (0,1,2)(0,1,0)

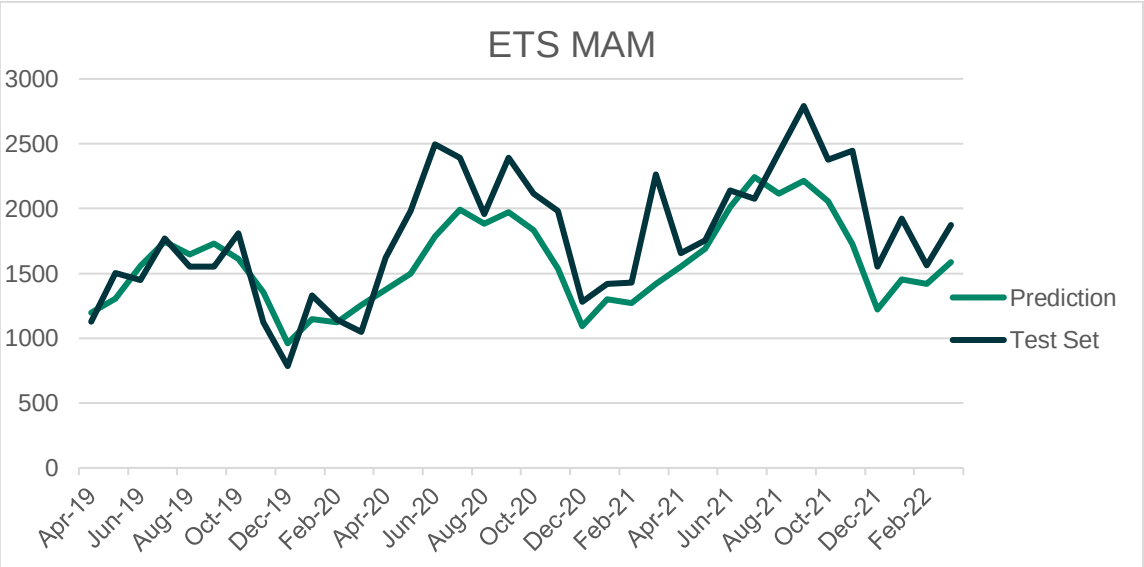


Figure 15. Verification of ETS MAM

The training set was then adjusted to contain data from April 2015-March 2021, with the test set from April 2021 to March 2022. In both the ARIMA and ETS MAM models, the total number of jobs predicted are then higher than the actual total number of jobs, with differences ranging from +14% to +15%.

	ARIMA	ETS MAM
Total Number of Jobs Predicted (21-22)	28,288	28,020
Difference to Historical Data (21-22)	+15%	+14%

Table 5. Train-test-split results (Train 7 yrs, Test 1yr)

The reason for the difference between these two train/test results is because the rate is increasing faster in more recent years, so if the model is trained on data to 2019 it is missing the most recent years data where the rate is increasing, hence the under-estimate. Whereas if data to 2021 is used for training, the job rate has started increasing at the end, so the next year is an over-estimate.

This also illustrates how sensitive the models are to any changes in the trend, given how short the testing period is.

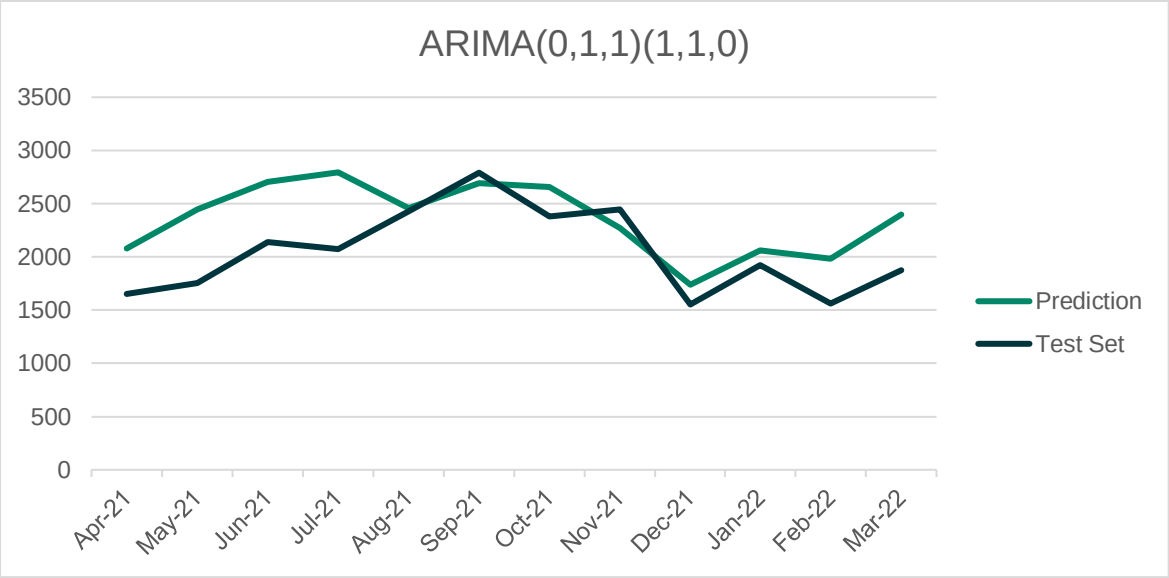


Figure 16. Verification of ARIMA(0,1,1)(1,1,0) 2021-2022

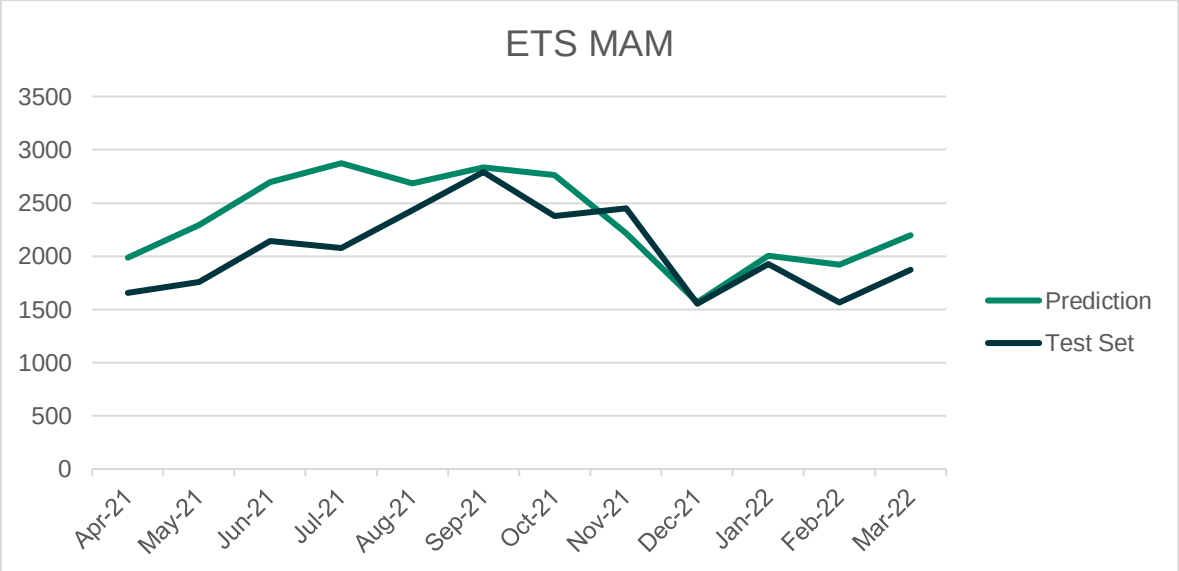


Figure 17. Verification of ETS MAM 2021-2022

4. Recommendation

Based on the MAPE value and increased seasonality observed, the predictions made using the ETS MAM model is determined to be the best fit to the observed data. This method is recognised as best suited to data with an underlying trend and with seasonality that increases over time, as it results in a curved forecast that also reproduces the seasonal changes in the data, as can be observed in Figure 18, matching the trend and increasing the seasonal variation with time, as per the input data.

Figure 18 shows the original AWS forecast plus the ETS MAM forecast, showing that the forecast is within the upper and lower bounds (as are all of the model forecasts described above.)

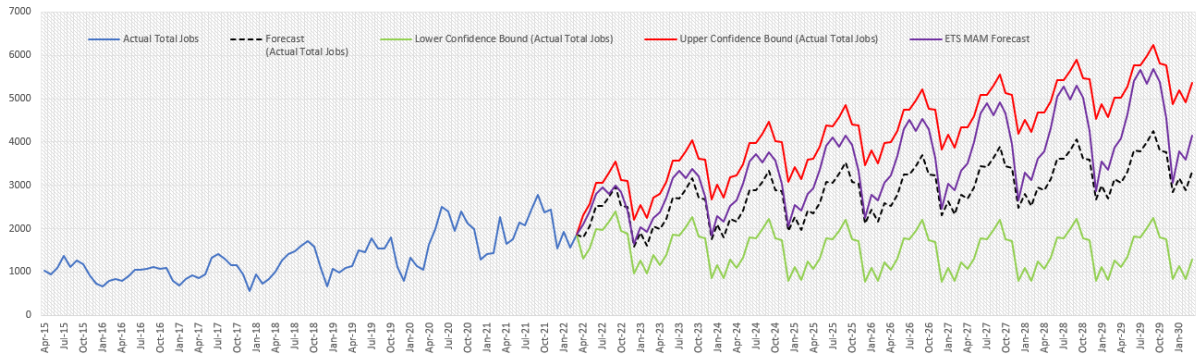


Figure 18. AWS original forecast with ETS MAM

It should be noted that these forecasts all assume the continuation of any existing maintenance regime.

Based on the “train-test-split” verification above, this model appears to be more sensitive to data from recent time periods. Therefore, it is recommended to re-run the analysis using the most recent data (i.e., to include 2022/23) to give a higher degree of confidence in the future forecasts.

Appendix A

A.1 Raw Job Count

Month	Actual Total Jobs
Apr-15	1,039
May-15	945
Jun-15	1,096
Jul-15	1,376
Aug-15	1,124
Sep-15	1,272
Oct-15	1,188
Nov-15	915
Dec-15	741
Jan-16	665
Feb-16	802
Mar-16	837
Apr-16	803
May-16	895
Jun-16	1,060
Jul-16	1,054
Aug-16	1,079
Sep-16	1,108
Oct-16	1,073
Nov-16	1,100
Dec-16	796
Jan-17	682
Feb-17	846
Mar-17	927
Apr-17	854
May-17	938
Jun-17	1,330
Jul-17	1,421
Aug-17	1,307
Sep-17	1,150
Oct-17	1,152
Nov-17	945
Dec-17	563
Jan-18	955
Feb-18	723
Mar-18	837

Month	Actual Total Jobs
Apr-18	1,018
May-18	1,258
Jun-18	1,406
Jul-18	1,483
Aug-18	1,603
Sep-18	1,711
Oct-18	1,587
Nov-18	1,100
Dec-18	664
Jan-19	1,083
Feb-19	984
Mar-19	1,097
Apr-19	1,129
May-19	1,505
Jun-19	1,449
Jul-19	1,772
Aug-19	1,550
Sep-19	1,552
Oct-19	1,809
Nov-19	1,123
Dec-19	785
Jan-20	1,330
Feb-20	1,142
Mar-20	1,050
Apr-20	1,623
May-20	1,981
Jun-20	2,497
Jul-20	2,394
Aug-20	1,958
Sep-20	2,393
Oct-20	2,116
Nov-20	1,981
Dec-20	1,281
Jan-21	1,418
Feb-21	1,429
Mar-21	2,263

Month	Actual Total Jobs
Apr-21	1,655
May-21	1,757
Jun-21	2,142
Jul-21	2,074
Aug-21	2,429
Sep-21	2,791
Oct-21	2,379
Nov-21	2,447
Dec-21	1,553
Jan-22	1,925
Feb-22	1,563
Mar-22	1,873

A.2 Estimated Job Count

Month	Excel ETS	R ETS AAA	ARIMA (0,1,1) (0,1,1) CSS	Holt Winters	R ETS MAM	ARIMA (0,1,1) (0,1,1) ML	R ETS AAA
01/04/2022	1,800	1,795	1,932	2,085	2,088	1,963	1,916
01/05/2022	2,050	2,042	2,139	2,303	2,387	2,166	2,127
01/06/2022	2,534	2,524	2,444	2,593	2,803	2,461	2,378
01/07/2022	2,521	2,509	2,470	2,610	2,951	2,496	2,425
01/08/2022	2,728	2,715	2,457	2,599	2,798	2,472	2,375
01/09/2022	2,979	2,964	2,696	2,793	2,997	2,693	2,502
01/10/2022	2,531	2,516	2,505	2,624	2,845	2,519	2,378
01/11/2022	2,498	2,482	2,322	2,415	2,425	2,324	2,095
01/12/2022	1,582	1,566	1,697	1,842	1,639	1,724	1,690
01/01/2023	1,906	1,889	2,004	2,137	2,037	2,023	1,905
01/02/2023	1,611	1,593	1,836	2,017	1,936	1,869	1,865
01/03/2023	2,055	2,036	2,151	2,320	2,249	2,166	2,036
01/04/2023	1,981	1,957	2,133	2,357	2,372	2,167	2,026
01/05/2023	2,231	2,204	2,339	2,575	2,707	2,370	2,237
01/06/2023	2,715	2,686	2,644	2,865	3,176	2,665	2,488
01/07/2023	2,702	2,671	2,670	2,882	3,338	2,700	2,535
01/08/2023	2,909	2,877	2,658	2,871	3,161	2,676	2,485
01/09/2023	3,160	3,126	2,896	3,065	3,382	2,897	2,611
01/10/2023	2,712	2,678	2,706	2,895	3,207	2,723	2,487
01/11/2023	2,679	2,644	2,523	2,687	2,730	2,528	2,205
01/12/2023	1,763	1,728	1,898	2,113	1,843	1,928	1,800
01/01/2024	2,087	2,051	2,204	2,409	2,288	2,227	2,014
01/02/2024	1,792	1,755	2,037	2,288	2,172	2,073	1,975
01/03/2024	2,236	2,198	2,352	2,592	2,521	2,371	2,146
01/04/2024	2,162	2,119	2,334	2,628	2,655	2,371	2,136
01/05/2024	2,412	2,367	2,540	2,846	3,027	2,574	2,346
01/06/2024	2,896	2,848	2,845	3,136	3,548	2,869	2,598
01/07/2024	2,883	2,834	2,871	3,153	3,726	2,904	2,645
01/08/2024	3,090	3,039	2,858	3,142	3,525	2,880	2,595
01/09/2024	3,341	3,289	3,097	3,336	3,767	3,101	2,721
01/10/2024	2,894	2,840	2,906	3,167	3,569	2,927	2,597
01/11/2024	2,860	2,807	2,723	2,958	3,035	2,732	2,315
01/12/2024	1,944	1,890	2,098	2,385	2,047	2,132	1,910
01/01/2025	2,268	2,214	2,405	2,680	2,539	2,431	2,124
01/02/2025	1,973	1,918	2,238	2,559	2,408	2,277	2,085
01/03/2025	2,417	2,360	2,553	2,863	2,793	2,575	2,255
01/04/2025	2,343	2,281	2,534	2,899	2,939	2,575	2,245
01/05/2025	2,593	2,529	2,740	3,117	3,348	2,778	2,456
01/06/2025	3,077	3,010	3,046	3,408	3,920	3,073	2,708
01/07/2025	3,064	2,996	3,071	3,424	4,114	3,108	2,754
01/08/2025	3,271	3,201	3,059	3,414	3,888	3,084	2,704

Month	Excel ETS	R ETS AAA	ARIMA (0,1,1) (0,1,1) CSS	Holt Winters	R ETS MAM	ARIMA (0,1,1) (0,1,1) ML	R ETS AAA
01/09/2025	3,522	3,451	3,297	3,607	4,152	3,305	2,831
01/10/2025	3,075	3,003	3,107	3,438	3,930	3,131	2,707
01/11/2025	3,041	2,969	2,924	3,230	3,340	2,936	2,425
01/12/2025	2,125	2,052	2,299	2,656	2,252	2,336	2,019
01/01/2026	2,450	2,376	2,605	2,952	2,790	2,635	2,234
01/02/2026	2,154	2,080	2,438	2,831	2,644	2,481	2,194
01/03/2026	2,598	2,522	2,753	3,135	3,064	2,779	2,365
01/04/2026	2,525	2,443	2,735	3,171	3,223	2,780	2,355
01/05/2026	2,775	2,691	2,941	3,389	3,668	2,982	2,566
01/06/2026	3,258	3,172	3,246	3,679	4,292	3,277	2,817
01/07/2026	3,246	3,158	3,272	3,696	4,501	3,312	2,864
01/08/2026	3,452	3,363	3,259	3,685	4,251	3,288	2,814
01/09/2026	3,703	3,613	3,498	3,879	4,537	3,509	2,941
01/10/2026	3,256	3,165	3,307	3,710	4,292	3,335	2,816
01/11/2026	3,223	3,131	3,124	3,501	3,645	3,140	2,534
01/12/2026	2,306	2,214	2,499	2,928	2,456	2,540	2,129
01/01/2027	2,631	2,538	2,806	3,223	3,041	2,839	2,344
01/02/2027	2,336	2,242	2,639	3,102	2,880	2,685	2,304
01/03/2027	2,779	2,684	2,954	3,406	3,336	2,983	2,475
01/04/2027	2,706	2,606	2,935	3,442	3,506	2,984	2,465
01/05/2027	2,956	2,853	3,142	3,660	3,989	3,186	2,675
01/06/2027	3,439	3,334	3,447	3,950	4,665	3,481	2,927
01/07/2027	3,427	3,320	3,472	3,967	4,889	3,517	2,974
01/08/2027	3,634	3,526	3,460	3,957	4,615	3,492	2,924
01/09/2027	3,884	3,775	3,698	4,150	4,923	3,714	3,050
01/10/2027	3,437	3,327	3,508	3,981	4,654	3,540	2,926
01/11/2027	3,404	3,293	3,325	3,772	3,950	3,344	2,644
01/12/2027	2,488	2,377	2,700	3,199	2,660	2,744	2,239
01/01/2028	2,812	2,700	3,006	3,494	3,292	3,043	2,453
01/02/2028	2,517	2,404	2,839	3,374	3,117	2,889	2,414
01/03/2028	2,960	2,847	3,154	3,678	3,608	3,187	2,584
01/04/2028	2,887	2,768	3,136	3,714	3,790	3,188	2,574
01/05/2028	3,137	3,015	3,342	3,932	4,309	3,391	2,785
01/06/2028	3,620	3,497	3,647	4,222	5,037	3,685	3,037
01/07/2028	3,608	3,482	3,673	4,239	5,276	3,721	3,084
01/08/2028	3,815	3,688	3,660	4,228	4,978	3,696	3,034
01/09/2028	4,065	3,937	3,899	4,422	5,308	3,918	3,160
01/10/2028	3,618	3,489	3,708	4,252	5,016	3,744	3,036
01/11/2028	3,585	3,455	3,526	4,044	4,255	3,548	2,754
01/12/2028	2,669	2,539	2,900	3,471	2,864	2,948	2,348
01/01/2029	2,993	2,862	3,207	3,766	3,543	3,247	2,563
01/02/2029	2,698	2,566	3,040	3,645	3,353	3,093	2,523
01/03/2029	3,141	3,009	3,355	3,949	3,879	3,391	2,694
01/04/2029	3,068	2,930	3,336	3,985	4,073	3,392	2,684

Month	Excel ETS	R ETS AAA	ARIMA (0,1,1) (0,1,1) CSS	Holt Winters	R ETS MAM	ARIMA (0,1,1) (0,1,1) ML	R ETS AAA
01/05/2029	3,318	3,178	3,543	4,203	4,630	3,595	2,895
01/06/2029	3,801	3,659	3,848	4,493	5,409	3,889	3,146
01/07/2029	3,789	3,645	3,873	4,510	5,664	3,925	3,193
01/08/2029	3,996	3,850	3,861	4,499	5,342	3,900	3,143
01/09/2029	4,247	4,100	4,100	4,693	5,693	4,122	3,270
01/10/2029	3,799	3,651	3,909	4,524	5,377	3,948	3,146
01/11/2029	3,766	3,618	3,726	4,315	4,560	3,752	2,863
01/12/2029	2,850	2,701	3,101	3,742	3,068	3,153	2,458
01/01/2030	3,174	3,025	3,407	4,037	3,794	3,451	2,673
01/02/2030	2,879	2,728	3,240	3,916	3,589	3,297	2,633
01/03/2030	3,322	3,171	3,555	4,220	4,151	3,595	2,804