Comparison of Different Forecasting Models on Global Nuclear Energy Production Data

Onder CATMABACAK,★
Physics and Astronomy Department, University of Oklahoma, 660 Parrington Oval, Norman, OK 73019

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
In this work, we claim to perform various forecast models to find the best fit to describe nuclear energy production by 2030 and 2050. The models that we used include simple to more complex methods. We have compared models different parameters and selected the best one to confront with other models for forecasting test data between 2021-2020. The outcome of each model was evaluated carefully and supported by statistical analyses. Results illustrate that our forecast models are in partial agreement with the 2021 nuclear power estimate report announced by International Atomic Energy Agency.

Key words: forecast – nuclear – energy

1 INTRODUCTION

Nuclear energy is the cleanest sustainable energy production method. Along with the increasing global energy demand, authorities have directed their interests toward atomic energy. Therefore, the primary concern of the predictions became the future of the energy industry, especially on both energy production and consumption.

Between 1964-1991, the Soviet Union has a series of nuclear tests under the name of *Project K nuclear tests* (3). Rival nations have joined the race, and the Soviet Union's series of nuclear tests and detonations have triggered commercial applications to make atomic energy a part of electricity production in the US and all around the globe. Nuclear technology starts to turn into a desire for all OECD countries due to a couple of reasons: It is clean, efficient, and produces reliable electricity on a large scale (7). Today, 10% of global electricity demand was granted by nuclear power plants, while this ratio goes up to 18% for OECD countries (4).

International Atomic Energy Agency annually publishes *"Energy, Electricity and Nuclear Power Estimates for the period up to 2050"* report. It tells us 16% increase by 2030 and 38% increase by 2050 in 2019 report (5) and 2.5% decrease by 2030 and going back to 2020 level by 2050 in 2021 report (6).

In this work, we performed a series of forecasts with statistical reasoning to establish the best model that can demonstrate the same outcome with the aforementioned report. The order of this paper was set from the simplest model to the most complex one for pedagogical purposes.

2 DATA

In this work, I used nuclear generation data set under the *"Statistical Review of World Energy – all data, 1965-2020"* on the British Petrol (BP) web page (1). The data consists of 56 yearly observations between 1965-2020, and there are no missing data or abnormalities. Variables are Year, Terawatt-hours, and Entity (World).

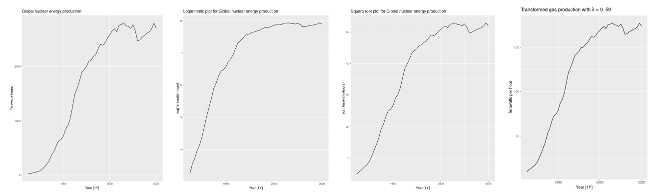

Figure 1. Global nuclear energy production between 1965-2020. Interesting points are the 2010 nuclear posture review and 2011 Fukushima-Daiichi nuclear power plant accident and 2019 Covid-19 outbreak. Plotted data (**top-left**), log plot (**top-right**), square root plot (**bottom-left**) and box-cox transformation (**bottom-right**)

We start with converting data into a time series object, *tsibble*, in R. The Entity and Year are the key and index parameters, respectively. The first results of plots are in Fig-1. In figure, we see four different plots; normal, logarithmic, square root and box-cox transformation ($\lambda = 0.59$). There are important events in our data that can affect our forecast. Soviet Union had a series of nuclear tests from 1964 to 1991 called Project K nuclear tests. Reflection of these tests, we monitor a stiff increase at the beginning of our data. In 2011, the power supply and three cooling units on reactors in the Fukushima-Daiichi nuclear power plant were strokes by a tsunami following a major earthquake (9.0 on the Richter scale). As a result, three reactors have lost their cooling systems and had nuclear meltdowns. There was one confirmed cancer death due to radiation exposure and 16 injured. However, this accident caused mass hysteria about nuclear energy. This scenario explains the sudden decrease in nuclear energy production in 2012. In Fig 1, we see substantial decrease in energy production. The second downward trend starts in 2019, the Covi-19 outbreak. However, these features have not been seen in the logarithmic plot, and there is not much change in trend for square root and box-cox transformations. Eventually, the data do not need any

★ E-mail: ondercatmabacak@ou.edu

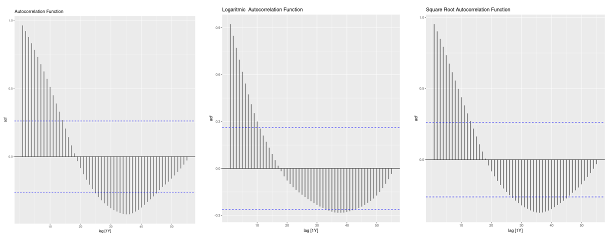

Figure 2. Auto-correlation functions of normal(**left**), logaritmic (**middle**) and square root (**right**) of nuclear energy production data. The correlograms show strong trend in residual data for the first 10 lags.

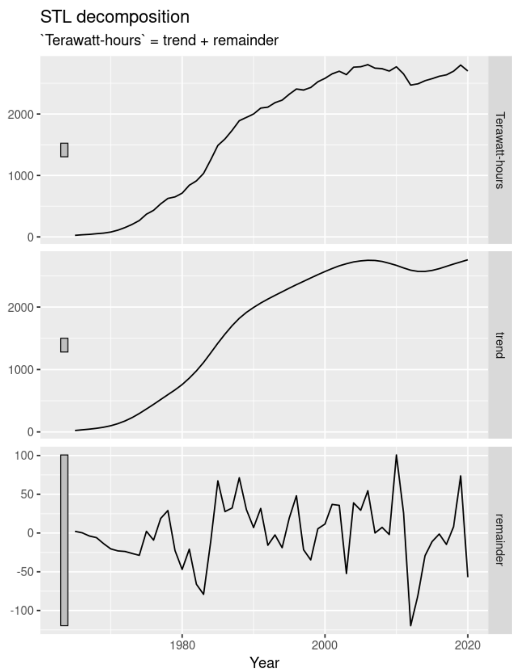

Figure 3. Seasonal and Trend decomposition using Loess method. Method indicates that there is no seasonality in the data.

transformation. Furthermore, in Fig-2, particular transformations do not provide better auto-correlation functions.

After discussing the influential events in the data, one can say that data is very straightforward, with no sign of cyclic or seasonal behavior. In addition to this, there is an upward trend with fluctuations at specific points in time.

3 METHOD

Simple forecasting methods such as Mean, Drift, Naive methods, and more complex methods like STL decomposition, ETS, ARIMA, Vector Auto-Regression, and Neural Network Auto-Regression, were used to perform our forecasts (2). We choose the best models for each method and combine them for a final prediction. Results are supported by statistical tests and analysis to find the best outcome from our forecasts.

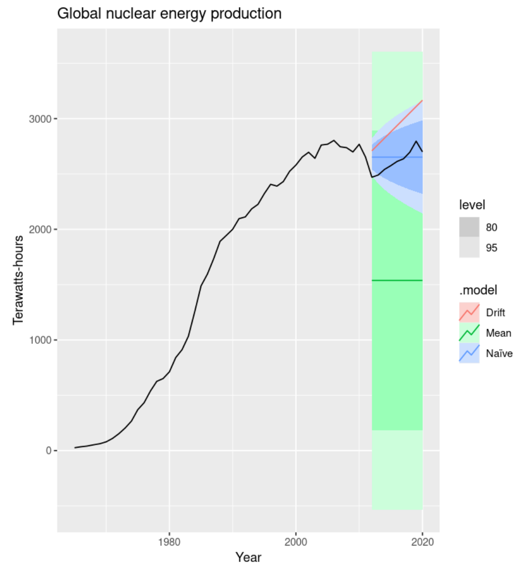

Figure 4. Forecasting test data (2012-2020) with training data (1965-2011). Three different models were applied: mean, drift, and naive.

.model	Entity	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Drift	World	Test	-325.	330.	325.	-12.4	12.4	4.65	3.82	0.175
Mean	World	Test	1076.	1081.	1076.	41.1	41.1	15.4	12.5	0.686
Naive	World	Test	-39.1	108.	91.3	-1.64	3.54	1.31	1.24	0.686

Figure 5. Statistics for mean, drift and naive models. The result indicates Naive model is the best fit among these models.

4 ANALYSIS AND DISCUSSION

4.1 STL Decomposition

We start our analysis with time series decomposition techniques.,We mentioned that our data do not have any seasonality in previous sections. Therefore, X11 and SEATS decomposition methods become unavailable for our data since these methods need seasonality in a data set. We performed STL decomposition in Fig-3 in which we see a clear upward trend. As we discussed in previous sections, there are evident varieties in the trend that are distinctive from fluctuations. The most influential ones are the 2011 Fukushima-Daiichi reactor accident and the 2019 Covid-19 pandemic. STL decomposition analysis will continue in later sections.

4.2 Mean, Drift and Naive methods

We conducted simple forecasting methods such as Mean, Drift, and Naive methods in Fig-4. Since our data have no seasonality, we skipped the seasonal Naive method. We create train data, between 1965-2011, for the models. That would be a subject of discussion to train the models before or after the accident. However, since our original purpose is to perform forecasts for 2030 and 2050, there is no point excluding the Fukushima-Daiichi observation in training data. We carry out the forecast for the next nine years, from 2012 to 2020.

As we interpret the forecast, the first thing that comes to our attention is that the Mean method has a massive prediction interval. Therefore, it is safe to say that we do not need statistical tests to eliminate the Mean method. The other two approaches, Drift and Naive, on the other hand, are more competitive. However, statistical

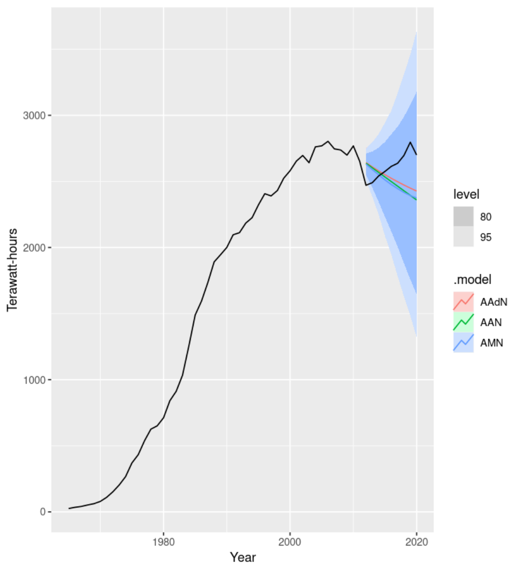

Figure 6. Exponential Smoothing (ETS) forecasts for different trends with additive error and no seasonal component. It has foreseen in all forecasts that, heavily affected by Fukushima-Daiichi accident, there is a downward trend in all models.

tests show that the Naive approach has an undisputed superiority over the Drift method. Among all three, only the Naive method was qualified to be used in final forecasts.

4.3 ETS models

Fig-6 displays forecast for various ETS models with additive error and no seasonality. We have three different trend options: additive damped additive, and multiplicative. Same rules hold for other ETS forecasts for multiplicative and automatically selected errors in Fig-7 and Fig-8.

In Fig-10 and Fig-9, `accuracy` and `glance` tests provide us all the information that we need to find optimum model. The first statistic test indicates that forecasts with automatically selected errors are indeed ETS(A, A, N), ETS(A, Ad, N), ETS(M, M, N), respectively. According to RMSE values, ETS(A, A, N)and ETS(A, Ad, N) have the best results. However, even though RMSE for ETS(A, Ad, N) is slightly better, ETS(A, A, N) has a better AICc score. On the contrary, ETS(M, M, N) model has better AICc than the other 2. However, the RMSE score does not exhibit the same performance. When it comes to choosing which statistical tests give a better solution, most of the time AICc is more advisable, however, in the final forecast, we cannot use AICc scores because of differencing on data in the ARIMA model makes it incomparable. Also ETS(A,X,N) models have smaller prediction interval in Fig-6. This would be another clue for choosing the correct model for final forecasts. Accordingly, we conclude that ETS(A, A, N) is the optimum ETS model for our data.

4.4 Non-seasonal ARIMA models

Before discussing the ARIMA model, check our data to see which model parameters we should use. In Fig-13, ACF and PACF plots are shown. The first ten lags, the first four lags in ACF, and the first lag in PACF contain meaningful information. We choose PACF since it has fewer lags beyond the white noise range. PACF plot (0,d,q) means we do not need any autoregression for the data. In Fig-12, a series

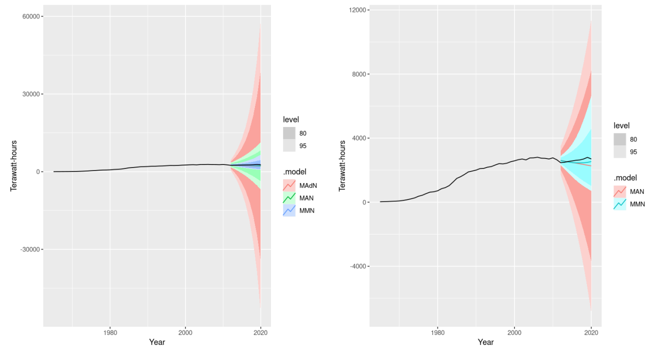

Figure 7. Exponential Smoothing (ETS) forecasts for different trends with multiplicative error and no seasonal component (**left**). Resolution is relatively low due to wide interval range of MAdN model. The same plot is also represented without "Ad" trend. (**right**).

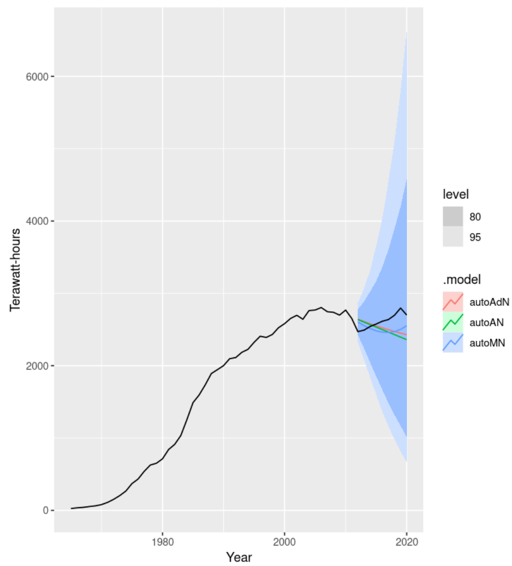

Figure 8. Exponential Smoothing (ETS) forecasts for different trends with automatically selected error and no seasonal component.

```
glance(ETS_A)
A tibble: 3 × 10
Entity .model sigma2 log_lik   AIC  AICc   BIC   MSE   AMSE   MAE
<chr>  <chr>   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl>
World  AAN     3175.   -278.  566.  567.  575. 2905.  8401.  39.2
World  AAdN    3223.   -278.  567.  569.  578. 2880.  8013.  40.5
World  AMN     3997.   -283.  577.  578.  586. 3657. 15903.  42.4
glance(ETS_M)
A tibble: 3 × 10
Entity .model  sigma2 log_lik   AIC  AICc   BIC    MSE   AMSE    MAE
<chr>  <chr>    <dbl>   <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl>  <dbl>
World  MAN    0.0287    -326.  662.  664.  672. 29809. 71550. 0.0647
World  MAdN   0.145     -360.  732.  734.  743.  3925. 10896. 0.110
World  MMN    0.00281   -269.  548.  550.  557.  3775. 16125. 0.0371
glance(ETS_auto)
A tibble: 3 × 10
Entity .model       sigma2 log_lik   AIC  AICc   BIC   MSE   AMSE     MAE
<chr>  <chr>         <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>   <dbl>
World  autoAN  3175.           -278.  566.  567.  575. 2905.  8401. 39.2
World  autoAdN 3223.           -278.  567.  569.  578. 2880.  8013. 40.5
World  autoMN      0.00281     -269.  548.  550.  557. 3775. 16125.  0.0371
```

Figure 9. Akaike information criterion (AIC), the corrected Akaike criterion (AICc) and the Bayesian information criterion (BIC), along with other parameters, for the ETS forecasts.

```
accuracy(ETS_A)
A tibble: 3 × 11
Entity .model .type          ME  RMSE   MAE   MPE  MAPE  MASE RMSSE      ACF1
<chr>  <chr>  <chr>       <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>     <dbl>
World  AAN    Training  -2.22  53.9  39.2  1.69  4.15 0.561 0.623  0.00583
World  AAdN   Training   4.53  53.7  40.5  4.29  8.71 0.581 0.620  0.00294
World  AMN    Training -15.8   60.5  42.4 -1.10  4.01 0.608 0.699 -0.0511
accuracy(ETS_M)
A tibble: 3 × 11
Entity .model .type          ME  RMSE   MAE   MPE  MAPE  MASE RMSSE     ACF1
<chr>  <chr>  <chr>       <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>    <dbl>
World  MAN    Training  21.5  173.    63.3 95.2  97.7  0.907 1.99  -0.0183
World  MAdN   Training   8.79  62.6  49.7  3.20 11.0  0.712 0.724 -0.416
World  MMN    Training -15.9   61.4  42.8 -1.20  3.83 0.614 0.710 -0.215
accuracy(ETS_auto)
A tibble: 3 × 11
Entity .model  .type          ME  RMSE   MAE   MPE  MAPE  MASE RMSSE      ACF1
<chr>  <chr>   <chr>       <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>     <dbl>
World  autoAN  Training  -2.22  53.9  39.2  1.69  4.15 0.561 0.623  0.00583
World  autoAdN Training   4.53  53.7  40.5  4.29  8.71 0.581 0.620  0.00294
World  autoMN  Training -15.9   61.4  42.8 -1.20  3.83 0.614 0.710 -0.215
```

Figure 10. Statistical tests for ETS forecasts.

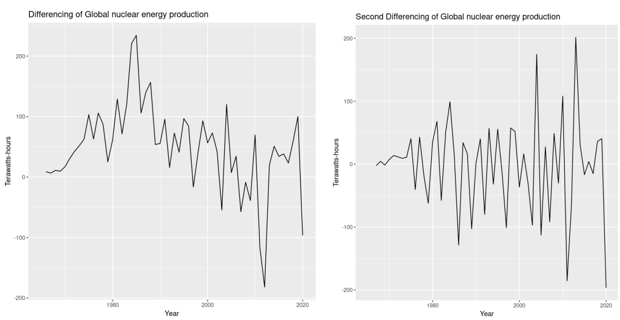

Figure 11. First (**left**) and second (**right**) differencing on global nuclear energy production data. The variance is fixed in the second plot.

```
nuc_tsbl %>%
  features(`Terawatt-hours`, unitroot_kpss)
A tibble: 1 × 3
Entity kpss_stat kpss_pvalue
<chr>      <dbl>       <dbl>
World       1.38        0.01
nuc_tsbl %>%
  features(difference(`Terawatt-hours`), unitroot_kpss)
A tibble: 1 × 3
Entity kpss_stat kpss_pvalue
<chr>      <dbl>       <dbl>
World      0.477      0.0468
nuc_tsbl %>%
  features(`Terawatt-hours`, unitroot_ndiffs)
A tibble: 1 × 2
Entity ndiffs
<chr>   <int>
World       2
```

Figure 12. Kwiatkowski–Phillips–Schmidt–Shin (KPSS) and ndiffs tests on the data. P-value suggest that we cannot reject the null hypothesis. The test tells us that second differencing is needed in order to stabilize the variance in the data.

of unit root tests provide valuable information about the number of differencing and p-value of the data. According to the tests, we failed to reject the null hypothesis, and residuals look like white noise, and the number of differencing processes that we need to apply to our data is 2. Hence, we can say data is not stationary and need a second differencing. The outcome of the first and second differencing can be seen in Fig-11.

Through the partial auto-correlation function and `ndiffs`, we concluded that our non-seasonal ARIMA models have constant auto-regression (p=0) and differencing (d=2) values. Now we check various combinations of moving average parameter, q. Fig 14-16 represent forecast of ARIMA models with various moving average values. Residual analyses, in Fig-17, demonstrate residuals, auto-correlation

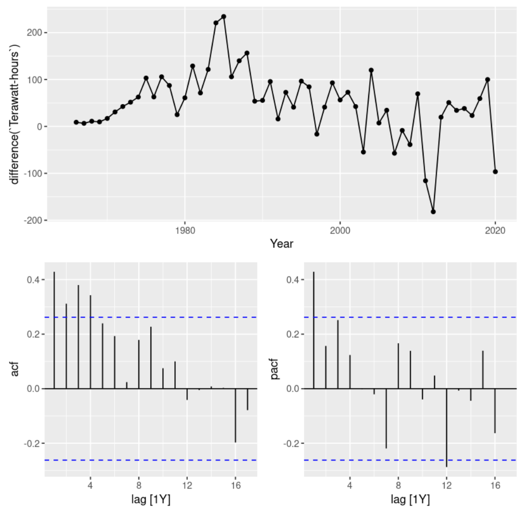

Figure 13. Auto-correlation and partial auto-correlation functions for global nuclear energy production data. Lags in auto-correlation functions indicate that PACF would be a better fit for our data.

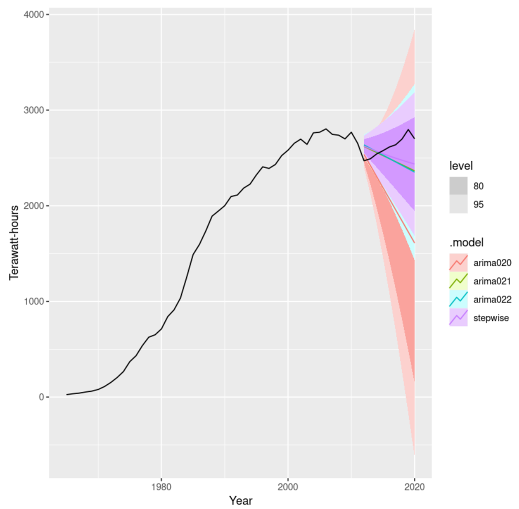

Figure 14. ARIMA forecasts of different moving averages + stepwise for PACF (p=0) and ndiffs =2.

function and histograms. For all non-seasonal ARIMA models, residuals look like white noise, except first lag in `arima020`. Likewise, only `arima020` models has an outlier on histogram. Histograms have mostly skewed to the right, except `arima022` which skewed to the left. Also, note that only `arima020` model has solely white noise in residual respect to the box-pierce and Ljung-box tests. Despite these favorable findings, the accuracy-test shows a different picture. We found that, in terms of residuals and p-value, `arima020` model has the only model that has pure white noise in residuals. However, it also has the worst `RMSE` and `AICc` scores. Hence, we have failed to choose `arima020` model as the best fit for our data.

We explore the most accurate prediction rather than interpretable/explainable one. In this respect, we directly look at RMSE since AICc is regarded for better balances complexity and accuracy. Our analysis indicates that `arima022` model have slightly better `RMSE` value, although `arima021` has a better `AICc`, it is not our priority to choose complexity - accuracy balance but the best ARIMA model.

```
accuracy(nuc_fit)
A tibble: 4 × 11
Entity .model   .type       ME  RMSE   MAE   MPE  MAPE  MASE RMSSE    ACF1
<chr>  <chr>    <chr>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>
World  arima020 Training -2.65  66.1  49.3 0.838  4.29 0.707 0.764 -0.461
World  arima021 Training -2.41  54.5  38.9 1.72   4.06 0.558 0.629 -0.0910
World  arima022 Training -2.20  53.9  39.1 1.68   4.08 0.561 0.623 -0.0101
World  stepwise Training  5.56  54.1  39.8 3.09   4.78 0.570 0.625 -0.0899
glance(nuc_fit)%>% arrange(AICc)
A tibble: 4 × 9
Entity .model   sigma2 log_lik   AIC  AICc   BIC ar_roots   ma_roots
<chr>  <chr>     <dbl>   <dbl> <dbl> <dbl> <dbl> <list>     <list>
World  arima021  3170.   -245.  494.  494.  498. <cpl [0]> <cpl [1]>
World  arima022  3176.   -245.  495.  496.  500. <cpl [0]> <cpl [2]>
World  stepwise  3127.   -250.  506.  506.  511. <cpl [1]> <cpl [1]>
World  arima020  4565.   -253.  509.  509.  511. <cpl [0]> <cpl [0]>
```

Figure 15. Statistics for ARIMA models.

```
nuc_fit%>%
  augment() %>%
  features(.innov, box_pierce, lag = 10, dof = 0)
A tibble: 4 × 4
Entity .model   bp_stat bp_pvalue
<chr>  <chr>      <dbl>     <dbl>
World  arima020   19.5     0.0349
World  arima021    6.18    0.800
World  arima022    5.17    0.879
World  stepwise    6.31    0.789
nuc_fit%>%
  augment() %>%
  features(.innov, ljung_box, lag = 10, dof = 0)
A tibble: 4 × 4
Entity .model   lb_stat lb_pvalue
<chr>  <chr>      <dbl>     <dbl>
World  arima020   22.0     0.0152
World  arima021    7.39    0.688
World  arima022    6.34    0.786
World  stepwise    7.53    0.675
```

Figure 16. Box-pierce and Ljung-box tests for different ARIMA models.

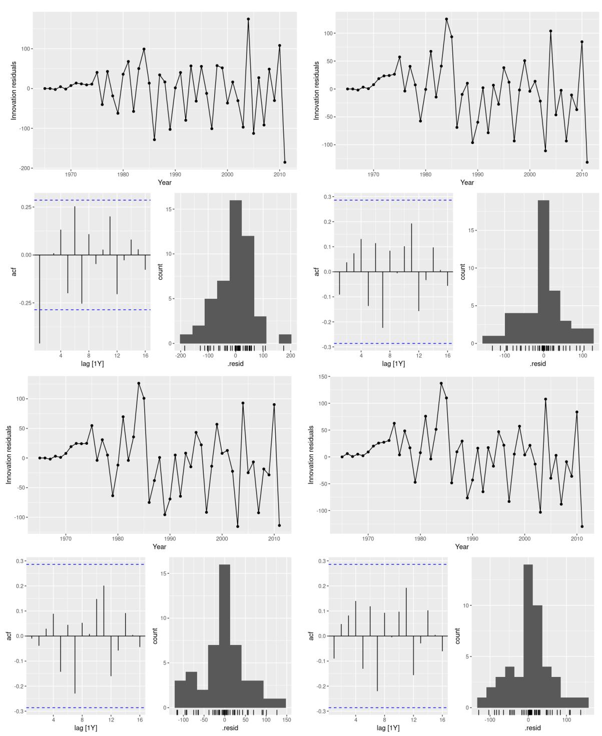

Figure 17. Residuals of ARIMA forecasts of different moving averages + stepwise for PACF (p=0) and ndiffs =2. Models: `arima020` (**top-left**), `arima021` (**top-right**), `arima022` (**bottom-left**), `stepwise` (**bottom-right**).

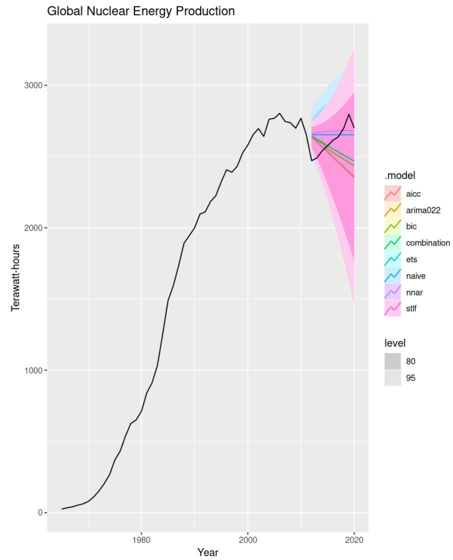

Figure 18. Combination of different forecast methods on test data (2012-2020).

```
nuc_models%>%
  augment() %>%
  features(.innov, box_pierce, lag = 10, dof = 0)
A tibble: 8 × 4
Entity .model      bp_stat bp_pvalue
<chr>  <chr>         <dbl>     <dbl>
World  aicc           6.62  0.761
World  arima022       5.17  0.879
World  bic            6.62  0.761
World  combination    5.19  0.878
World  ets            5.13  0.882
World  naive         30.5   0.000708
World  nnar           9.22  0.511
World  stlf           5.13  0.882
nuc_models%>%
  augment() %>%
  features(.innov, ljung_box, lag = 10, dof = 0)
A tibble: 8 × 4
Entity .model      lb_stat lb_pvalue
<chr>  <chr>         <dbl>     <dbl>
World  aicc           8.25  0.604
World  arima022       6.34  0.786
World  bic            8.25  0.604
World  combination    6.47  0.774
World  ets            6.29  0.790
World  naive         34.0   0.000187
World  nnar          11.1   0.350
World  stlf           6.29  0.790
```

Figure 19. Box pierce and Ljung box tests of different forecast methods on test data (2012-2020).

```
.model      Entity .type    ME  RMSE   MAE   MPE  MAPE  MASE RMSSE  ACF1
<chr>       <chr>  <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
aicc        World  Test   79.7  183.  155.   2.82  5.86  2.23  2.11 0.709
arima022    World  Test  120.   224.  186.   4.32  6.99  2.67  2.59 0.708
bic         World  Test   79.7  183.  155.   2.82  5.86  2.23  2.11 0.709
combination World  Test   58.5  166.  139.   2.01  5.26  1.99  1.92 0.708
ets         World  Test  115.   220.  183.   4.14  6.87  2.62  2.54 0.708
naive       World  Test  -39.1  108.   91.3 -1.64  3.54  1.31  1.24 0.686
nnar        World  Test  -60.3  112.   94.1 -2.45  3.67  1.35  1.29 0.683
stlf        World  Test  115.   220.  183.   4.14  6.87  2.62  2.54 0.708
```

Figure 20. Statistics of different forecast methods on test data (2012-2020).

4.5 Forecast Combination

Throughout our inquiries, we found out that `STL`, `Naive`, `ETS(A,A,N)`, `ARIMA(0,2,2)` are the best among in individual tests with verious models. In addition to these models, we also added vector auto-regression (`aicc` and `bic`), neural network auto-regression (`nnar`) and average of all these models (`combination`).

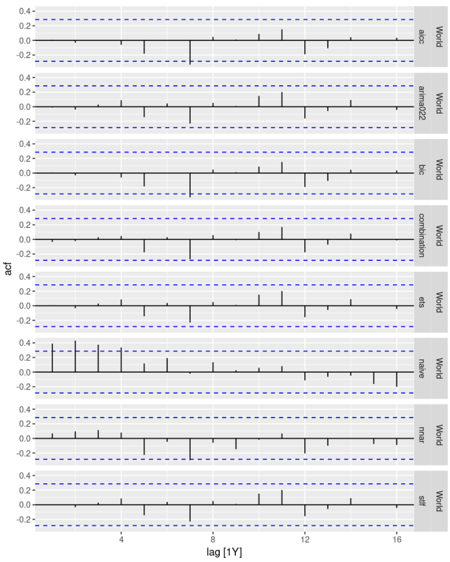

Figure 21. Residuals of different forecast methods on test data (2012-2020).

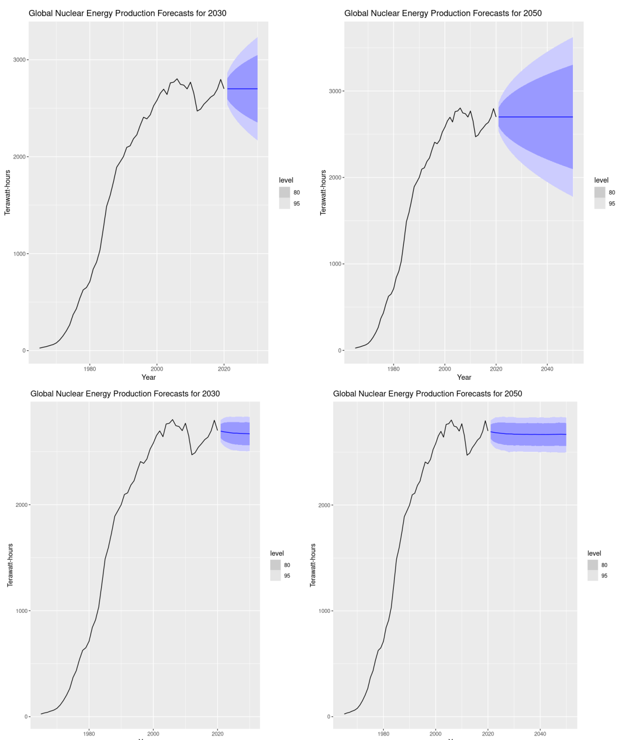

Figure 22. Forecasts by naive (*top*) and neural network auto-regression (**bottom**) methods for 2030 and 2050.

As we performed previous models, we start with training our models with train data in Fig-18. Therefore, we will continue with actual forecasts for 2030 and 2050. As we stated in the introduction, *"Energy, Electricity and Nuclear Power Estimates for the period up 2050"* report of International Atomic Energy Agency tells us 16% increase by 2030 and 38% increase by 2050 in 2019 report and 2.5% decrease by 2030 and going back to 2020 level by 2050 in 2021 report.

Auto-correlation functions for multiple models are in in Fig-21. All the models, except `naive`, seem like have residuals consist of white noise. However, Fig-19 indicates that the only p-value that fails to reject null hypothesis belongs to `naive` approach. Therefore, only residuals of naive methods look like white noise. In Fig-20, we see multiple accuracy tests. Naive method has the best scores in all tests except `ME` and `ACF1`. Runner up, neural network auto-regression, has slightly poorer score.

5 RESULTS

Previously we provided plots and statistics for training data and determined that the Naive method (`naive`) model is the best fit for our data. Therefore, in Fig-22, we applied Naive method (`naive`) and neureal network auto-regression (`nnar`) to the data to predict the nuclear energy production by 2030 and 2050. Frankly, `naive` model does not predict 2.5% decrease in energy production for 2030 forecast while it is the case on `nnar` forecast. Likewise, `naive` foretells the same level with 2020 but without predicting little decrease on 2050 forecast. On the other hand, `nnar` model predicts the decrease on 2030 forecsat but does not show the same level with 2020 on 2050 forecast. Lastly, `nnar` model represents Covid-19 related decrease in energy production more realistically on its prediction interval.

6 CONCLUSION

We performed our analyses to find the best forecast model among selected methods. As 14th-century logician and theologian William of Ockham stated, sometimes simplicity is better than complexity (Occam's razor). In our case, the naive method verifies that the simplest solution is indeed the best one. In our case, the naive approach justifies that the simplest method is the best fit for our data. Therefore, we estimate how nuclear energy generation has changed by 2030 and 2050. The importance of this study is to provide some perspective about the future of the cleanest and most sustainable energy production method through comparative analysis. Our examination leads us to the conclusion that neural network auto-regression is the best method that reflects the results of the International Atomic Energy Agency. Our results are in partial agreement with the annual reports from 2021. The 2019 report was published pre-pandemic time. Therefore, predictions do not echo reality anymore.

REFERENCES

https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html

Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 25 Oct 2021.

https://en.wikipedia.org/wiki/Soviet_Project_K_nuclear_tests

https://world-nuclear-news.org/Articles/IAEA-forecasts-doubling-of-nuclear-capacity-by-205

INTERNATIONAL ATOMIC ENERGY AGENCY, Energy, Electricity and Nuclear Power Estimates for the Period up to 2050, Reference Data Series, 2019

INTERNATIONAL ATOMIC ENERGY AGENCY, Energy, Electricity and Nuclear Power Estimates for the Period up to 2050, Reference Data Series, 2021

The Future of Nuclear Power AN INTERDISCIPLINARY MIT STUDY, 2003, ISBN 0-615-12420-8

This paper has been typeset from a TEX/LATEX file prepared by the author.