Time Series Forecasting of Freshwater Withdrawals in Germany Using ARIMA and Prophet

1. Introduction

This section aims to forecast freshwater withdrawals in Germany using time series analysis. Freshwater withdrawals are a key indicator of national water demand and are closely linked to sustainability metrics such as GDP and water stress. Two forecasting models are employed: ARIMA, a classical statistical method, and Prophet, a modern additive time series model developed by Facebook. The goal is to evaluate their performance and generate reliable forecasts.

2. R Libraries used

Library	Purpose
forecast	Classical ARIMA modeling
prophet	Decomposable time series forecasting with automatic changepoint detection
tidyverse	Data wrangling and visualization
lubridate	Date parsing and manipulation
tseries	Time series stationarity testing (e.g., Augmented Dickey-Fuller Test)

3. Variable and Country Selection Rationale

3.1. Target Variable: Freshwater Withdrawals

Freshwater withdrawals were chosen as the target variable because they represent total water demand across all economic sectors. This variable is directly involved in the calculation of other key indicators: GDP (USD) is computed as the product of water productivity and withdrawals, while water stress is derived from the ratio of withdrawals to renewable water resources, multiplied by 100. These mathematical relationships highlight the centrality of withdrawals in understanding both economic performance and environmental pressure. Forecasting this variable therefore supports broader insights into sustainability and development trends.

3.2. Target Country: Germany

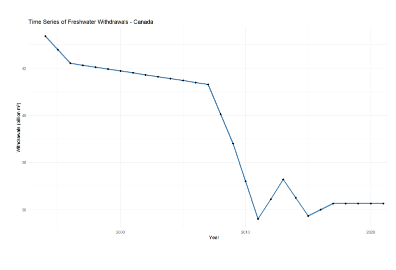
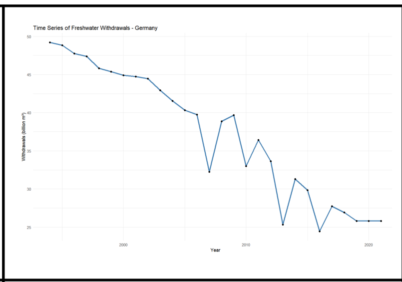
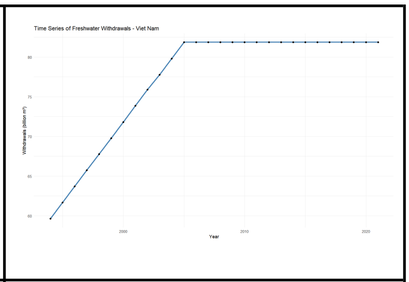
		
Canada(“CAN”)	Germany(“DEU”)	Vietnam(“VNM”)

Figure 1. Time Series of Freshwater Withdrawals of various countries

Germany was selected after reviewing the freshwater withdrawal patterns of numerous countries. Its time series exhibits a clear and consistent long-term downward trend, with moderate year-to-year variability that provides structure without excessive noise. Also, there are no long flat segments, which often limit model learning, and the values remain within a stable, realistic range, avoiding outliers or sudden regime shifts. In contrast, many countries showed less favorable patterns—some, like Vietnam, had time series that became completely flat in recent years, offering no meaningful variation to model. Others, such as Canada, displayed irregular drops and prolonged stagnation, which reduced forecastability. Therefore, Germany proved to be the most suitable candidate for time series forecasting.

4. Forecasting Framework

As previously mentioned, both the AutoRegressive Integrated Moving Average (ARIMA) model and the Facebook Prophet model were applied to model the historical freshwater withdrawals in Germany. Conceptual explanations of these models are provided in Appendix D.

4.1. ARIMA Forecasting

The forecasting procedure followed a two-step approach. First, the model was trained on the full available time span to produce a five-year forecast for future withdrawals. This was used primarily for visualization and exploratory insight. Then, two separate evaluation methods were applied to assess model performance. The first was a traditional holdout approach: the dataset was split into a training period (up to 2017) and a test period (2018–2021). A new ARIMA model was fitted on the training data, and its predictive accuracy was evaluated on the test set using MAE, RMSE, and MAPE. The second method was a rolling forecast evaluation, in which the model was repeatedly trained on an expanding window and used to predict the following year. This rolling process, spanning from 2003 to 2020, enabled a more comprehensive view of the model's generalization over time.

```r
# Convert to time series object
ts_germany <- ts(df$Value, start = min(df$Year), frequency = 1)
# frequency = 1 means yearly data (as opposed to monthly = 12, quarterly = 4).

# Estimate lambda for Box-Cox transformation
lambda <- BoxCox.lambda(ts_germany)
cat("Suggested Box-Cox lambda:", lambda, "\n")

# Check for stationarity using Augmented Dickey-Fuller Test
adf_test <- tseries::adf.test(ts_germany, alternative = "stationary")
print(adf_test)

# Plot ACF and PACF to visually inspect autocorrelation
acf(ts_germany, main = "ACF - Germany Withdrawals")
pacf(ts_germany, main = "PACF - Germany Withdrawals")

# Fit ARIMA model: This uses the auto.arima() function to automatically determine
# the best-fitting ARIMA model for the time series.
arima_model <- auto.arima(ts_germany)
```

Figure 2. Code snippet – ARIMA model design and forecasting logic.

The time series was explicitly defined with frequency = 1 to indicate annual observations, distinguishing it from monthly or quarterly data where seasonal components would be more relevant. The auto.arima() function from the forecast package was used to automatically select the best-fitting ARIMA model by minimizing the corrected Akaike Information Criterion (AICc). This process identifies the optimal combination of autoregressive (AR), differencing (I), and moving average (MA) terms without requiring manual specification.

```
> lambda <- BoxCox.lambda(ts_germany)
> cat("Suggested Box-Cox lambda:", lambda, "\n")
Suggested Box-Cox lambda: 1.999924
```

Figure 3. Code snippet - Box-Cox lambda estimation for variance stability

Before the analysis was conducted, three diagnostic tests—the Box-Cox lambda estimation for variance stability, the Augmented Dickey-Fuller (ADF) test for stationarity, and the autocorrelation function (ACF/PACF) plots for autocorrelation structure —were performed to verify the assumptions required for ARIMA modeling. First, the Box-Cox lambda was estimated to evaluate whether a variance-stabilizing transformation was necessary. The result indicated a lambda of approximately 2.0, suggesting a square transformation. However, since the data exhibited no clear signs of heteroscedasticity and interpretability would be compromised, no transformation was applied.

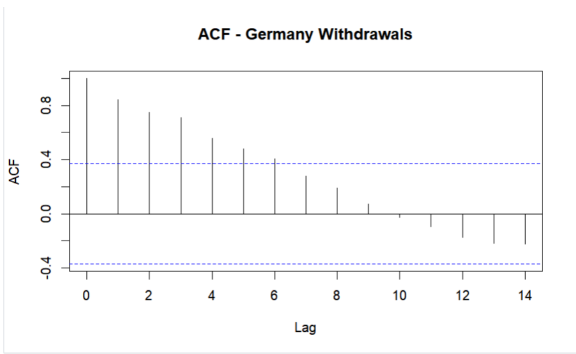
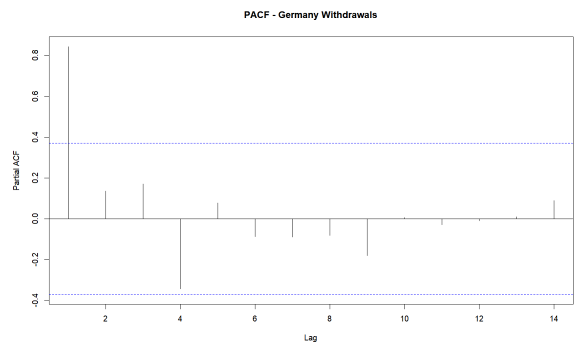
```
> adf_test <- tseries::adf.test(ts_germany, alternative = "stationary")
> print(adf_test)

        Augmented Dickey-Fuller Test

data:  ts_germany
Dickey-Fuller = -2.6909, Lag order = 3, p-value = 0.3082
alternative hypothesis: stationary
```

Figure 4. Code snippet - Augmented Dickey-Fuller (ADF) stationarity test applied to the original time series

ARIMA models assume that the underlying time series is stationary, meaning its statistical properties (mean, variance, autocorrelation) do not change over time. The resulting p-value of 0.3082 indicated non-stationarity. However, since the auto.arima() function automatically identifies and applies the appropriate differencing required for stationarity, no manual differencing was performed.

<table>
<tr>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>ACF plot of Germany's freshwater withdrawals.</td>
<td>PACF plot of Germany's freshwater withdrawals.</td>
</tr>
</table>

The autocorrelation function (ACF) plot shows significant positive autocorrelations at multiple lags, particularly from lag 1 to lag 6, all exceeding the 95% confidence bounds (blue dashed lines). This pattern indicates strong temporal dependence, suggesting that past values heavily influence future values. The gradual decline is characteristic of a non-stationary series, supporting the need for differencing before ARIMA modeling(which will be done by auto.arima()).
The partial autocorrelation function (PACF) plot shows a strong spike at lag 1, followed by a sharp cutoff with all values within the confidence bounds. This pattern is typical of an autoregressive process of order 1 (AR(1)), suggesting that only the immediate previous value contributes significantly to the current value when controlling for intermediate lags. This structure further supports the use of an ARIMA model with an autoregressive component.

4.2 Prophet Forecasting

As with ARIMA, Prophet was first fitted on the full dataset to produce a five-year forecast for visualization. Then, two evaluation methods were applied: one model was trained on data up to 2017 and tested on the 2018–2021 period using MAE, RMSE, and MAPE; another was evaluated using a rolling forecast approach, where the model was repeatedly retrained on expanding windows to forecast the following year. This consistent two-step approach enabled a direct and fair comparison between the Prophet and ARIMA models across both static and dynamic evaluation settings.

```r
# Prepare data for Prophet
prophet_df <- df %>%
  rename(ds = Year, y = Value) %>%
  mutate(ds = ymd(paste0(ds, "-01-01")))

# Fit Prophet model with yearly seasonality disabled (not needed for annual data)
prophet_model <- prophet(prophet_df, yearly.seasonality = FALSE)

# Create future dates (5 years ahead)
future <- make_future_dataframe(prophet_model, periods = 5, freq = "year")

# Forecast future withdrawals
forecast_prophet <- predict(prophet_model, future)
```

Figure 5. Code snippet – Prophet model setup and forecasting logic.

Prophet is an additive time series model that decomposes the data into trend, seasonality, and holiday effects, and is particularly effective for handling irregular trends and missing values. In this case, only the trend component was modeled, as the data is annual and does not exhibit seasonal patterns. The input data was formatted to match Prophet's required structure, with the time variable renamed to ds and the target variable to y.

5. Forecasting Results and Evaluation

5.1. ARIMA Forecasting

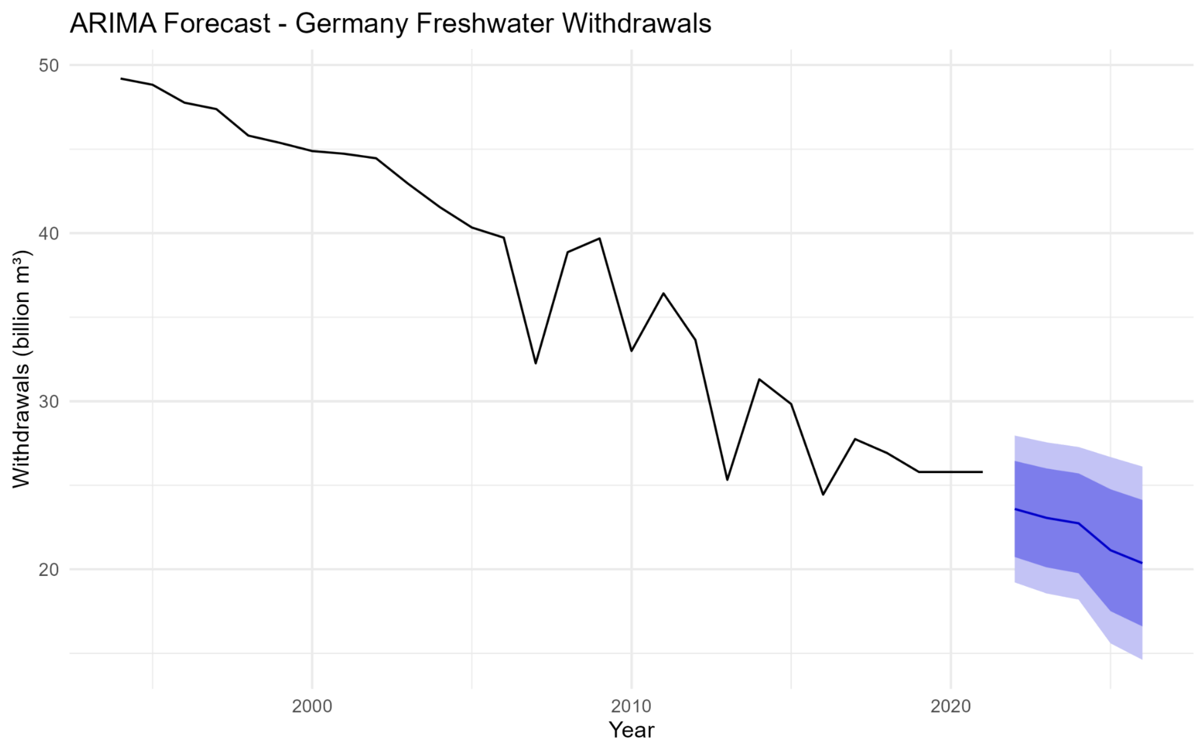


Figure 6. ARIMA Forecast of Germany's Freshwater Withdrawals (2022–2026)

This plot illustrates the five-year forecast generated by the ARIMA model trained on the full historical dataset. The black line represents the observed data up to 2021, while the blue line extends the forecasted trend through 2026. The shaded areas denote the 80% and 95%

confidence intervals, capturing increasing uncertainty further into the forecast horizon. The model projects a continued decline in freshwater withdrawals, consistent with the long-term trend observed in the past data.

	A	B
1	Year	Forecast
2	2022	23.59159
3	2023	23.05934
4	2024	22.73899
5	2025	21.13809
6	2026	20.36587

Figure 7. Forecasted Values from the ARIMA Model (2022–2026)

This table lists the exact forecasted values corresponding to the ARIMA plot shown earlier, which provides a numerical reference. Together with the visual forecast in Figure 4, these results offer a practical basis for future decision-making and water resource policy planning.

```
> summary(arima_model)
Series: ts_germany
ARIMA(2,1,0) with drift

Coefficients:
          ar1      ar2    drift
      -0.7582  -0.6711  -0.9060
s.e.   0.1335   0.1259   0.1721

sigma^2 = 4.968:  log likelihood = -59.08
AIC=126.15   AICc=127.97   BIC=131.34

Training set error measures:
                     ME     RMSE      MAE        MPE     MAPE      MASE     ACF1
Training set 0.01225276 2.063632 1.283608 -0.2315765 3.962897 0.5439712 0.114933
```

Figure 8. ARIMA Model Summary

The final model selected by auto.arima() was an ARIMA(2,1,0) with drift, indicating that the time series was differenced once and modeled using two autoregressive terms. This aligns with the results of the stationarity test, which confirmed the need for differencing. While the PACF plot of the original series suggested a strong AR(1) structure, the automated model selection process identified that including a second autoregressive term improved the overall model fit based on information criteria - possibly due to residual autocorrelation not fully captured by a single lag. The inclusion of drift captures the underlying trend in the differenced series.

All estimated coefficients, including the drift term (−0.9060), were statistically significant, as their absolute values exceeded approximately twice their respective standard errors—a common rule of thumb when formal p-values are not reported. The negative drift reflects a consistent downward trend in the data.

Model selection was based on the corrected Akaike Information Criterion (AICc), which is the default criterion. The selected model yielded AIC = 126.15, AICc = 127.97, and BIC = 131.34. On the full dataset, the model achieved a MAPE of 3.96%, indicating a strong in-sample fit. The residual autocorrelation at lag 1 (ACF1 = 0.11) was low, suggesting that most temporal patterns were successfully captured. While MAPE is a widely used metric, it can be sensitive to low actual values; however, in this case, the scale and stability of the data mitigate that concern.

```
> checkresiduals(arima_model)  # Includes Ljung-Box test, residual ACF, histogram

	Ljung-Box test

data:  Residuals from ARIMA(2,1,0) with drift
Q* = 6.8181, df = 4, p-value = 0.1458

Model df: 2.   Total lags used: 6
```

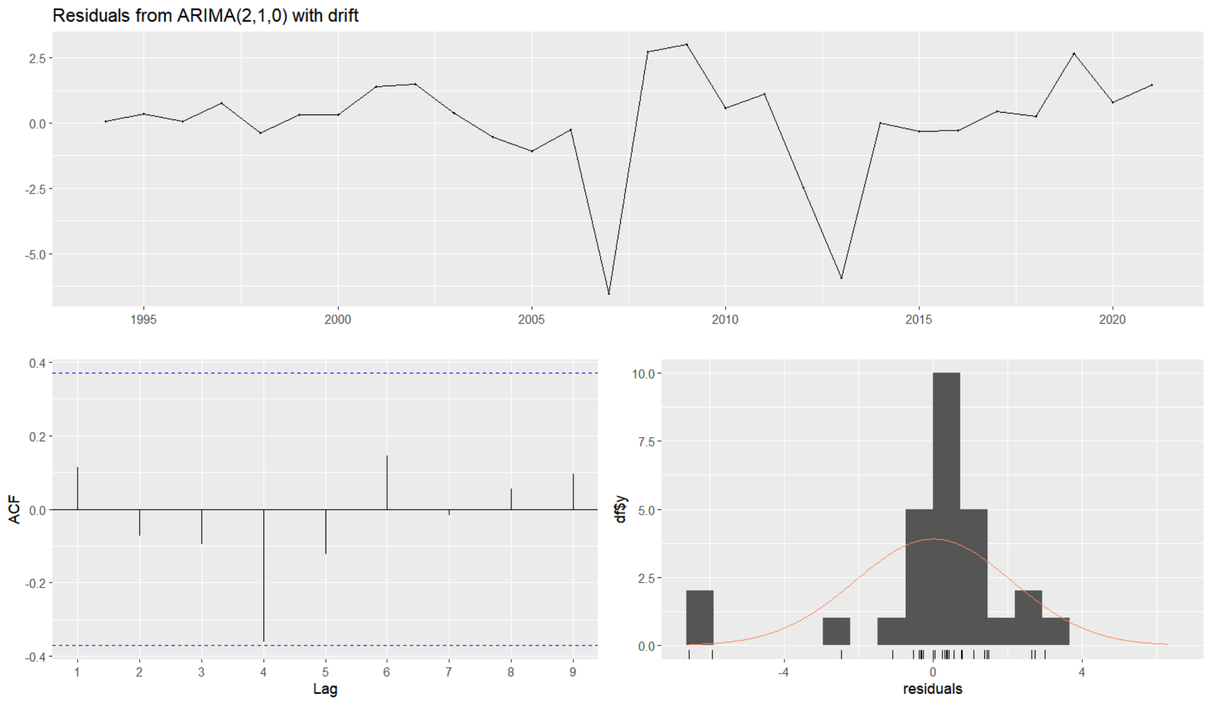


Figure 9, 10. Residual diagnostics for the ARIMA(2,1,0) with drift model

The residual diagnostics show no major violations of model assumptions. In the time series plot (top panel), the residuals appear randomly scattered around zero, with no visible trend or seasonality—suggesting no systematic error remains. In the autocorrelation function (ACF) plot, all lags fall within the 95% confidence bounds, which is a common threshold for determining statistical significance. The histogram of residuals is roughly bell-shaped, supporting the assumption of normality, although slight skewness is observed. Most importantly, the Ljung-Box test returned a p-value of 0.1458 (above the 0.05 significance level), indicating that there is insufficient evidence of residual autocorrelation. Based on these criteria, the model is considered a statistically valid fit for the data.

5.2. Prophet Forecasting

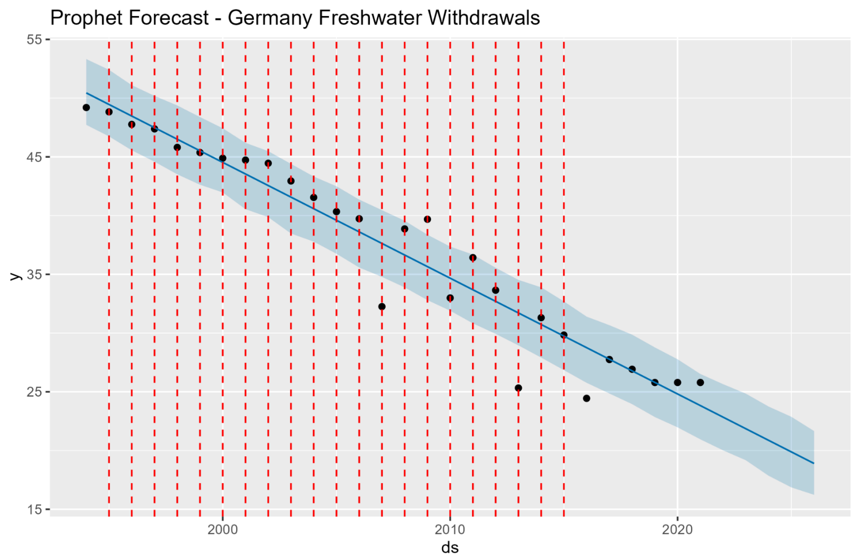


Figure 11. Prophet Forecast of Germany's Freshwater Withdrawals (2022–2026)

This plot presents the five-year forecast generated by the Prophet model trained on the full dataset. The blue line represents the model's predicted trend, while the shaded region shows the forecast uncertainty interval. The black dots correspond to the actual historical observations, most of which fall within the confidence band—indicating a good model fit. The vertical red dashed lines mark the changepoints automatically detected by Prophet, where the trend is allowed to shift. Although no abrupt structural breaks are visible in the series, these changepoints allow the model to adapt flexibly to subtle changes in trend. Overall, Prophet forecasts a continued decline in freshwater withdrawals, consistent with the observed long-term downward trajectory.

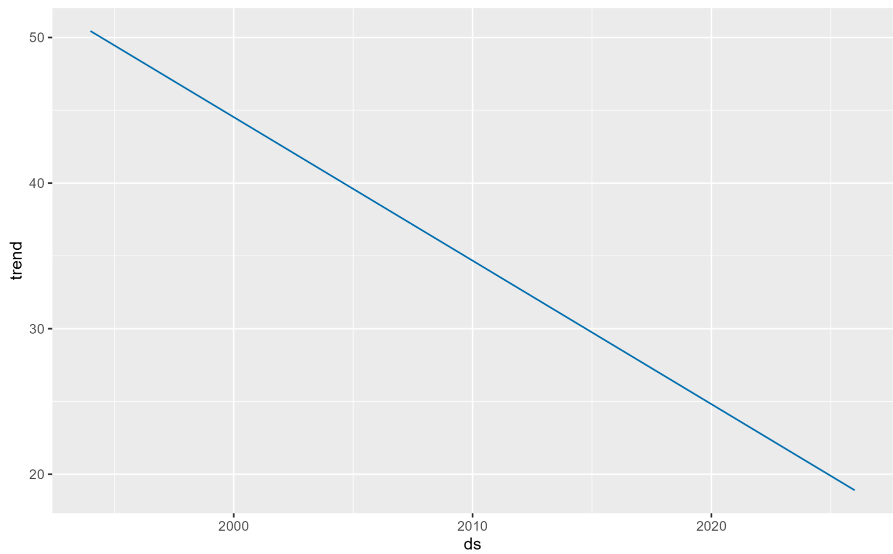


Figure 12. Prophet Forecast Components – Germany's Freshwater Withdrawals (2022–2026)

Note: Since the seasonality and holiday effect are absent, the line in the graph is identical to that of *Figure 7*.

#######	22.84169	20.01963	25.64703
#######	21.85656	19.17608	24.84236
#######	20.87142	17.8599	23.74378
#######	19.88358	16.86555	22.86875
#######	18.89845	16.23397	21.65674

Figure 13. Forecasted and Fitted Values from the Prophet Model

This table contains the Prophet model's output across the full time span. The yhat column contains fitted values for historical years (1994–2021) and forecasted values for future years (2022–2026), with yhat_lower and yhat_upper indicating uncertainty bounds. (Only the five forecasted values are displayed in the figure due to image size concerns.)

5.3. Model Performance Evaluation and Comparison

As previously mentioned, both ARIMA and Prophet were first trained on data up to 2017 and evaluated on a held-out test set spanning 2018 to 2021 to compare model performance. Forecast accuracy was measured using three standard metrics: MAE, RMSE, and MAPE. The results for each model are summarized below.

<table>
<tr><td><pre>> cat("ARIMA Metrics:\n")
ARIMA Metrics:
> cat("MAE: ", MAE, "\nRMSE:", RMSE,
"\nMAPE:", MAPE, "%\n")
MAE: 2.011173
RMSE: 2.262085
MAPE: 7.778954 %</pre></td><td><pre>> cat("Prophet Metrics:\n")
Prophet Metrics:
> cat("MAE: ", MAE, "\nRMSE:", RMSE,
"\nMAPE:", MAPE, "%\n")
MAE: 1.457615
RMSE: 1.676682
MAPE: 5.620193 %</pre></td></tr>
<tr><td>ARIMA</td><td>Prophet</td></tr>
</table>

Metric	Description
MAE (Mean Absolute Error)	Average magnitude of the forecast errors, regardless of direction. Simple and easy to interpret. However, does not penalize large errors more than small ones.
RMSE (Root Mean Squared Error)	Heavily penalizes large errors, making it useful when large deviations are especially undesirable. However, less interpretable than MAE due to squaring and square-rooting.
MAPE (Mean Absolute Percentage Error)	Average absolute error expressed as a percentage of the actual values. Intuitive and unit-free, making it easy to compare across different scales. However, can be misleading when actual values are very small or zero (division by small numbers inflates error).

On the held-out test set (2018–2021), the Prophet model slightly outperformed ARIMA across all three error metrics. Prophet achieved a lower MAE (1.46 vs. 2.01), RMSE (1.68 vs. 2.26), and MAPE (5.62% vs. 7.78%). While these results might suggest that Prophet provides a better in-sample fit on the test window, model performance can vary depending on the time segment evaluated. Therefore, to obtain a more robust and generalizable assessment, we proceeded with a rolling forecast evaluation over a longer time horizon.

```
# Define rolling parameters
start_year <- 2003
end_year <- 2020
horizon <- 1          # Forecast 1 year ahead in each loop
```

Figure 14. Rolling forecast parameter settings used for evaluation

The rolling forecast was configured with a start year of 2003 and an end year of 2020, using a 1-year forecast horizon. This means the model was repeatedly trained on data up to each year from 2003 to 2020 and used to predict the value for the following year. This setup produced 18 evaluation rounds, enabling a more reliable estimate of model performance over time.

<table>
<tr>
<td><pre>> cat("ARIMA Rolling Forecast
Metrics:\n")
ARIMA Rolling Forecast Metrics:
> cat("Mean MAE: ",
mean(rolling_mae), "\n")
Mean MAE: 2.726816
> cat("Mean RMSE:",
mean(rolling_rmse), "\n")
Mean RMSE: 2.726816
> cat("Mean MAPE:",
mean(rolling_mape), "%\n")
Mean MAPE: 8.651565 %</pre></td>
<td><pre>> # Print average results
> cat("Prophet Rolling Forecast
Metrics:\n")
Prophet Rolling Forecast Metrics:
> cat("Mean MAE: ",
mean(rolling_mae_prophet), "\n")
Mean MAE: 2.160427
> cat("Mean RMSE:",
mean(rolling_rmse_prophet), "\n")
Mean RMSE: 2.160427
> cat("Mean MAPE:",
mean(rolling_mape_prophet), "%\n")
Mean MAPE: 7.199758 %</pre></td>
</tr>
<tr>
<td>ARIMA</td>
<td>Prophet</td>
</tr>
</table>

The results from the rolling forecast evaluation indicate that Prophet achieved slightly lower average error metrics compared to ARIMA across all measures. Specifically, Prophet recorded a mean MAE of 2.16 and a mean MAPE of 7.20%, while ARIMA showed higher errors, with a mean MAE of 2.73 and MAPE of 8.65%. The RMSE values were equal to the MAEs in both models, likely due to the relatively stable error distribution.
This pattern aligns with the earlier held-out test results, where Prophet also showed a modest advantage over ARIMA. A possible reason for this outcome—at least in the context of the present dataset—is Prophet's use of piecewise linear trend fitting and automatic changepoint detection, which may have helped it adapt more flexibly to subtle structural changes in the declining trend. This characteristic could have contributed to its slightly improved accuracy in this particular case. However, it should be noted that in cases where the underlying time series exhibits stable linear trends without significant structural changes,

ARIMA may provide more robust and interpretable results than Prophet, due to its reliance on well-established statistical assumptions and simpler model structure.
While the overall differences in performance are not dramatic, both models produced relatively low average error rates, suggesting that they are reasonably effective forecasters for this type of national-level water withdrawal data. That said, the slightly lower and more consistent errors observed with Prophet across both evaluation settings provide tentative support for its use in similar long-term forecasting tasks.

6. Interpretation of Forecast Results and Conditional Policy Recommendations

The forecasting results for Germany’s freshwater withdrawals, generated by both ARIMA and Prophet models, suggest a continued downward trend in the near future. By 2026, ARIMA projects withdrawals to decline to approximately 20.4 billion m^3, while Prophet offers a similar estimate of 18.9 billion m^3 — both down from roughly 26 billion m^3 in 2021. While the models demonstrate strong statistical performance, it is important to recognize that such forecasts may reflect a range of underlying factors. A declining withdrawal trend could indicate positive developments, such as improved water-use efficiency or structural economic shifts, but it might also result from less favorable conditions, including economic downturns or climate-related constraints. Interpreting these results in context is therefore essential, particularly when deriving policy recommendations.

6.1. Potential Drivers Behind the Forecasted Withdrawal Trend

Several plausible scenarios may explain the observed and predicted reduction in withdrawals:

Cause Type	Possible Explanation
Efficiency Gains	Improved irrigation systems, industrial reuse, household conservation.
Economic Restructuring	Shift toward less water-intensive sectors such as services.
Economic Decline	Contraction in manufacturing or agriculture reducing demand.
Climate Impact	Droughts or groundwater depletion limiting actual withdrawal capacity.
Data Artefacts	Changes in monitoring methods or definitions over time.

Each scenario carries very different policy implications. For example, efficiency gains may justify infrastructure optimization, whereas climate-driven reductions may require emergency preparedness and adaptation investment.
To clarify which possibilities may be driving the observed trend, further analysis should be conducted using exploratory data analysis (EDA), linear regression, and potentially classification techniques. These methods help uncover patterns and predictors that explain the trajectory of withdrawals more clearly and inform targeted policy interventions.

6.2. Scenario-Based Policy Recommendations

The table below outlines appropriate responses based on the dominant driver of the withdrawal trend:

Scenario	Policy Direction
Efficiency improvements	Expand support for innovation, precision agriculture, smart water systems.
Sectoral shift	Align infrastructure plans with new economic structures and long-term demand.
Economic downturn	Avoid irreversible cuts in supply capacity; adopt flexible, modular planning.
Climate-related decline	Strengthen drought resilience, water storage, and non-traditional supply methods.
Measurement changes	Improve data transparency, harmonize indicators, and monitor consistency.

These recommendations are not mutually exclusive; multiple causes may operate simultaneously. Therefore, it is desirable that Policymakers adopt adaptive frameworks that can respond dynamically as new information emerges.

7. Conclusion

This study has presented a time series analysis of Germany's freshwater withdrawals using two forecasting approaches: ARIMA and Prophet. In both the held-out test and rolling forecast evaluations, Prophet slightly outperformed ARIMA across all error metrics. This pattern, observed in this specific dataset, may suggest Prophet's advantage in capturing non-linear trends and adapting to structural changes, particularly through its automatic changepoint detection. While the difference in performance was modest, both models proved to be suitable for forecasting long-term national-level water usage trends.

Both models projected a continued downward trend in withdrawals through 2026. While these projections are statistically reliable, their real-world interpretation requires caution. A declining trend may reflect favorable developments such as improved efficiency or structural economic shifts, but it could also signal more concerning drivers such as economic slowdown or climate-related constraints. The time series models capture the "what," but not the "why."

To address this limitation, further analysis should be conducted using exploratory data analysis, regression analysis, and potentially classification techniques. These tools will help clarify the underlying causes of withdrawal changes and provide a more nuanced understanding of the observed patterns.

Ultimately, this study lays the foundation for a broader research framework that combines accurate forecasting with causal interpretation. By integrating time series forecasting with complementary analytical methods, this research aims to support more informed, adaptive, and sustainable water resource planning — both within Germany and in other countries facing similar water management challenges.