Forecasting Groundwater Availability in Auser Aquifer, Italy

ENV 797 Final Project

Emma Kaufman and Jenn McNeill

2024-04-10

Introduction, Motivation, Relevance, Objectives

This project focuses on predicting water availability in an Italian aquifer managed by Acea Group, a leading Italian utility operator. The group provides water services to 9 million inhabitants across Italy.

In order to best service their customers, the Acea Group must understand how much water is available in the water bodies from which they extract. Forecasting water availability in a water body is necessary to ensure daily consumption needs are met.

In 2022, the UN reported that groundwater provides "half of the volume of water withdrawn for domestic use by the global population" and that water use is expected to grow 1% per year over the next thirty years (UN, 2022). Thus, it is important to explore how much groundwater will be available in the future. If groundwater levels are forecasted to drop dramatically, these models can be used to urge for more efficient water use practices and investments into groundwater recharge strategies.

This study is focused in Italy due to the comprehensive data available in the region. The objective of the following report is to examine the best models and exogenous variables that can help accurately predict groundwater levels, in the hope that these findings offer insight about how to focus groundwater forecasting efforts in other regions across the globe.

Dataset information

The data for this analysis were collected from the Acea Group Smart Water Analytics Competition on Kaggle hyperlink. As a utility operator, they are concerned with preserving their water bodies which include a combination of water springs, lakes, rivers, and aquifers. The nine unique datasets from this kaggle competition each had different attributes and characteristics. This report focuses time series modeling and forecasting on the Auser Aquifer. The goal is to predict the amount of water in the Auser Aquifer by modeling depth to groundwater and simultaneously evaluating how rainfall and temperature may impact our predictions as exogenous variables.

The dataset for the Auser Aquifer includes daily depth to groundwater measurements (in meters) from five different wells across the north and south sectors. Wells SAL, PAG, CoS, and DIEC represent the northern unconfined portion of the aquifer, while Well LT2 represents the southern confined portion of the aquifer. The dataset also contains daily temperature data at four sites, daily rainfall data at ten sites, and daily volume data from five different water treatment facilities that extract water from this aquifer. A sample of these data are shown in Tables 1 and 2, and the units for the measurements are found in Table 3. GitHub Repository Linkhyperlink.

Table 1: Example Auser Aquifer Data Rows 1-5

Date	05-17-11	05-18-11	05-19-11	05-20-11	05-21-11
Rainfall_Gallicano	0.0	0.0	0.0	0.8	0.0
Rainfall_Pontetetto	0	0	0	0	0
Rainfall_Monte_Serra	0	0	0	0	0
Rainfall_Orentano	0.0	0.0	0.0	0.2	0.0
Rainfall_Borgo_a_Mozzano	0	0	0	0	0
Rainfall_Piaggione	0	0	0	0	0
Rainfall_Calavorno	0	0	0	0	0
Rainfall_Croce_Arcana	0.0	0.0	0.0	0.0	1.0
Rainfall_Tereglio_Coreglia_Antelminelli	0.0	0.0	0.0	0.6	0.0
Rainfall_Fabbriche_di_Vallico	0.0	0.0	0.0	0.4	0.0
Depth_to_Groundwater_LT2	-12.97	-12.93	-12.92	-12.93	-12.92
Depth_to_Groundwater_SAL	-5.92	-5.93	-5.95	-5.95	-5.95
Depth_to_Groundwater_PAG	-2.34	-2.46	-2.41	-2.48	-2.43
Depth_to_Groundwater_CoS	-6.22	-6.27	-6.32	-6.39	-6.49
Depth_to_Groundwater_DIEC	-3.79	-3.80	-3.80	-3.81	-3.81
Temperature_Orentano	16.05	17.20	19.25	20.65	20.40
Temperature_Monte_Serra	12.80	15.25	15.35	16.40	17.60
Temperature_Ponte_a_Moriano	17.20	19.00	19.95	20.15	21.35
Temperature_Lucca_Orto_Botanico	17.45	19.00	20.10	21.60	21.15
Volume_POL	-9936.0	-9936.0	-9936.0	-9936.0	-9936.0
Volume_CC1	-16377.12	-16377.12	-16377.12	-16377.12	-16377.12
Volume_CC2	-12823.49	-12823.49	-12823.49	-12823.49	-12823.49
Volume_CSA	0	0	0	0	0
Volume_CSAL	0	0	0	0	0
Hydrometry_Monte_S_Quirico	0.17	0.18	0.16	0.15	0.15
Hydrometry_Piaggione	-1.04	-1.04	-1.04	-1.04	-1.05

Table 2: Example Auser Aquifer Data Head Rows 6-10

Date	05-22-11	05-23-11	05-24-11	05-25-11	05-26-11
Rainfall_Gallicano	0.0	0.0	0.0	0.0	0.0
Rainfall_Pontetetto	0	0	0	0	0
Rainfall_Monte_Serra	0	0	0	0	0
Rainfall_Orentano	0.0	0.0	0.0	0.0	0.0
Rainfall_Borgo_a_Mozzano	0	0	0	0	0
Rainfall_Piaggione	0	0	0	0	0
Rainfall_Calavorno	2	0	0	0	0
Rainfall_Croce_Arcana	0.2	0.0	0.0	0.0	0.0
Rainfall_Tereglio_Coreglia_Antelminelli	0.0	0.0	0.0	0.0	0.0
Rainfall_Fabbriche_di_Vallico	0.0	0.0	0.0	0.0	0.0
Depth_to_Groundwater_LT2	-12.91	-12.93	-12.94	-12.94	-12.93
Depth_to_Groundwater_SAL	-5.97	-6.01	-6.03	-6.05	-6.09
Depth_to_Groundwater_PAG	-2.54	-2.46	-2.47	-2.59	-2.61
Depth_to_Groundwater_CoS	-6.62	-6.70	-6.70	-6.72	-6.72
Depth_to_Groundwater_DIEC	-3.82	-3.83	-3.84	-3.84	-3.85
Temperature_Orentano	21.65	22.15	24.35	23.30	23.85
Temperature_Monte_Serra	18.65	20.25	20.20	21.30	20.25
Temperature_Ponte_a_Moriano	22.60	23.70	24.30	24.95	24.25
Temperature_Lucca_Orto_Botanico	22.55	23.60	24.05	24.60	24.70
Volume_POL	-9439.2	-9936.0	-9936.0	-9936.0	-9936.0
Volume_CC1	-15558.26	-16377.12	-16377.12	-16377.12	-16377.12
Volume_CC2	-12182.31	-12823.49	-12823.49	-12823.49	-12823.49
Volume_CSA	0	0	0	0	0
Volume_CSAL	0	0	0	0	0
Hydrometry_Monte_S_Quirico	0.15	0.15	0.14	0.15	0.14
Hydrometry_Piaggione	-1.05	-1.05	-1.05	-1.05	-1.06

Table 3: Acea Group Auser Aquifer Data Structure

Variables	Units
Date	Date
Rainfall_Gallicano	Millimeters
Rainfall_Pontetetto	
Rainfall_Monte_Serra	
Rainfall_Orentano	
Rainfall_Borgo_a_Mozzano	
Rainfall_Piaggione	
Rainfall_Calavorno	
Rainfall_Croce_Arcana	
Rainfall_Tereglio_Coreglia_Antelminelli	
Rainfall_Fabbriche_di_Vallico	
Depth_to_Groundwater_LT2	Meters
Depth_to_Groundwater_SAL	
Depth_to_Groundwater_PAG	
Depth_to_Groundwater_CoS	
Depth_to_Groundwater_DIEC	
Temperature_Orentano	Celcius
Temperature_Monte_Serra	
Temperature_Ponte_a_Moriano	
Temperature_Lucca_Orto_Botanico	
Volume_POL	Cubic Meters
Volume_CC1	
Volume_CC2	
Volume_CSA	
Volume_CSAL	
Hydrometry_Monte_S_Quirico	Meters
Hydrometry_Piaggione	

Data Wrangling

Each well's depth to groundwater data were converted into time series objects after extraneous zero values towards the end of the dataset were converted into NAs. Then the tsclean() function was run to fill in the gaps of missing data with interpolated values. The resulting cleaned time series are displayed in Figure 1 below. Additionally, the daily depth measurements were averaged by month to create a monthly time series. Temperature and rainfall were also averaged by month to be used as exogenous variables in models with these monthly time series.

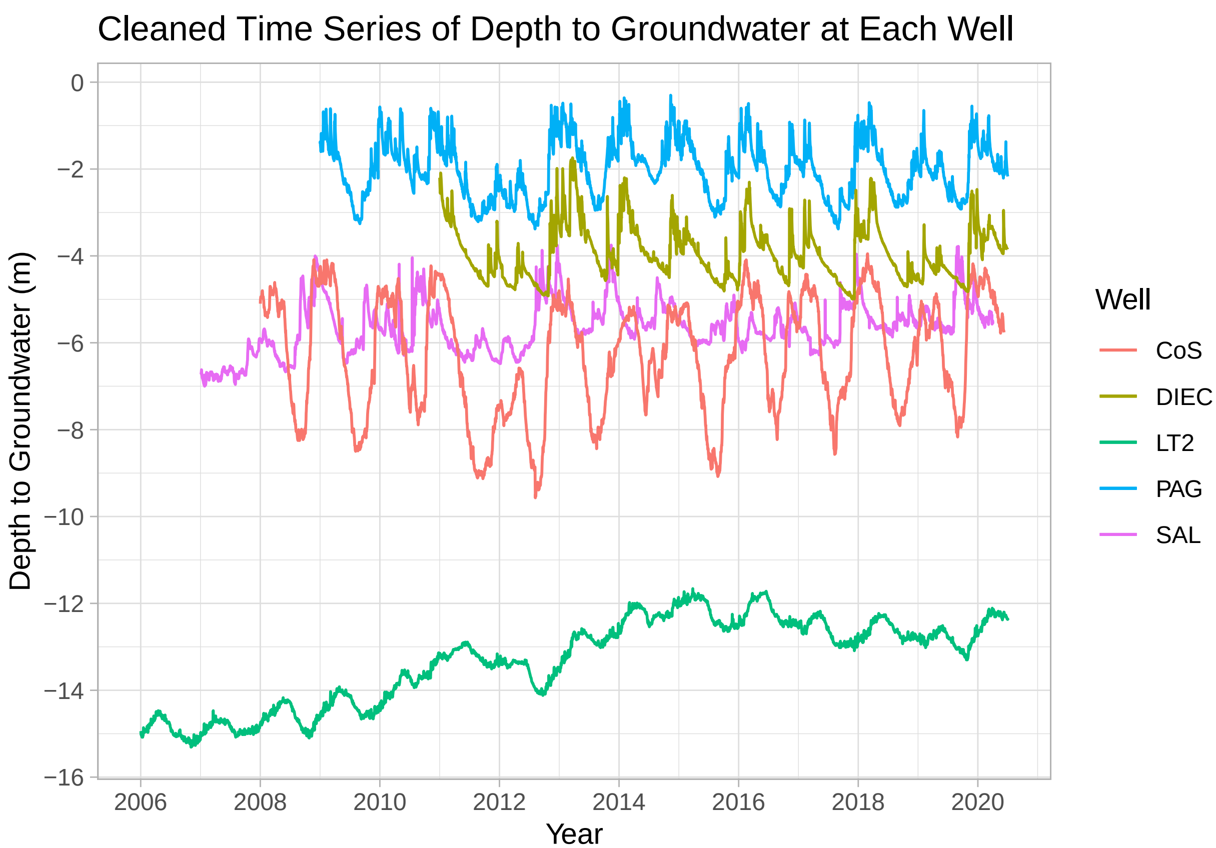


Figure 1: Depth to Groundwater Time Series

Analysis (Methods and Models)

The correlation function was run on depth to groundwater data to discern whether the depth to groundwater values at the five wells were correlated to one another. The results are displayed in Figure 2. The depth to groundwater of the four north wells were well correlated, while the southern well was weakly correlated to the others. As a result, the rest of the analysis takes a closer look at one northern well (SAL) within the unconfined region of the aquifer, and the one southern well (LT2) within the confined region of the aquifer. These two wells also had the longest time series data available, as displayed in Figure 3.

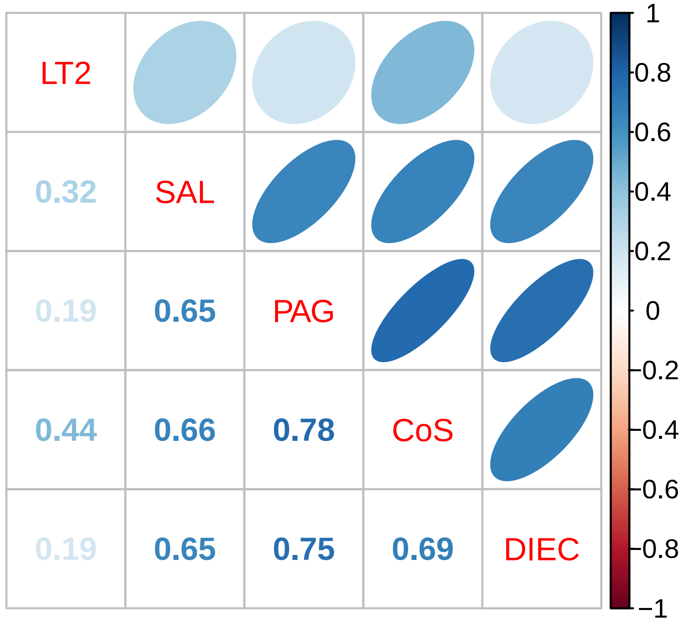


Figure 2: Correlation of Well Depths in Auser Aquifer

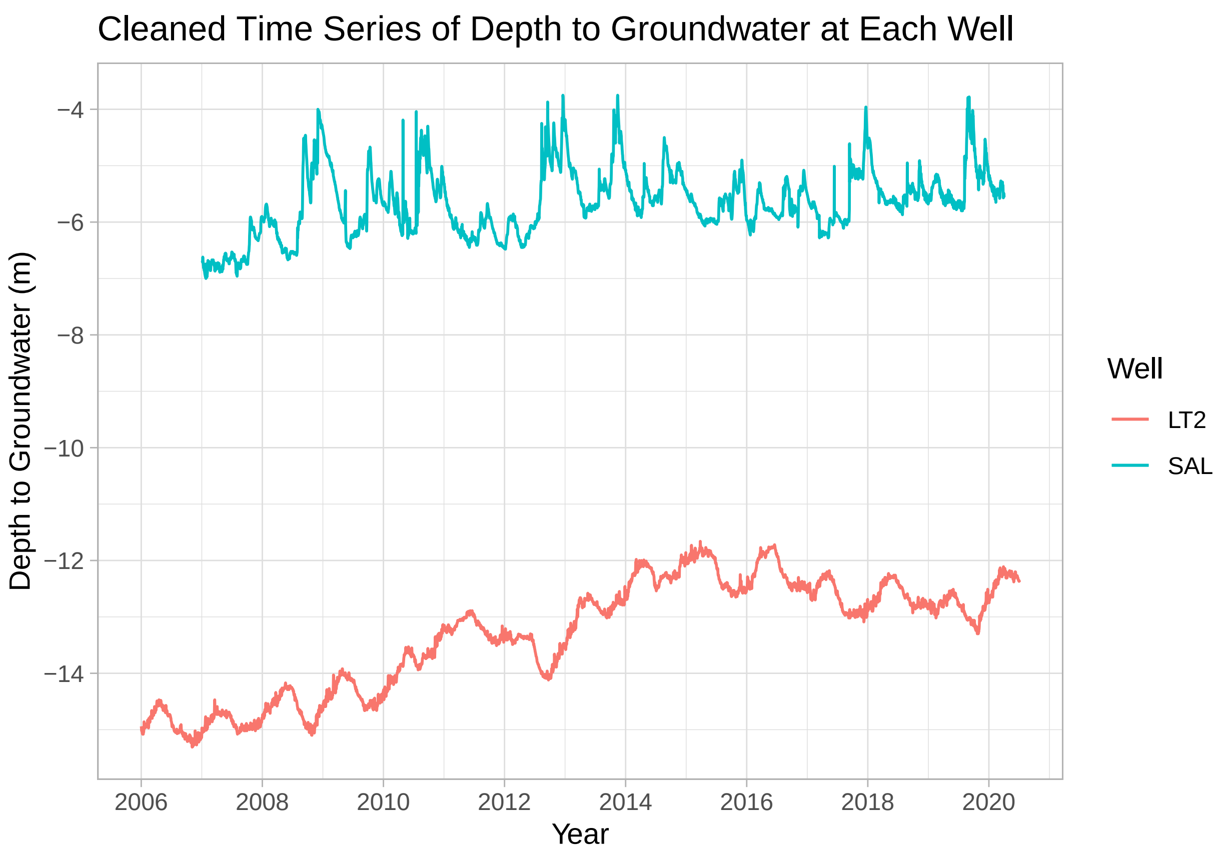


Figure 3: LT2 and SAL Time Series

The ACF and PACF for each well were then plotted using a lag time of five years to understand the seasonal pattern of the data; these results are displayed in Figure 4. Both ACF graphs display peaks and troughs at regular intervals, which is indicative of seasonality. Knowing that temperature and rainfall affect aquifer storage, it makes sense that the depth to groundwater in the aquifer changes with respect to the season.

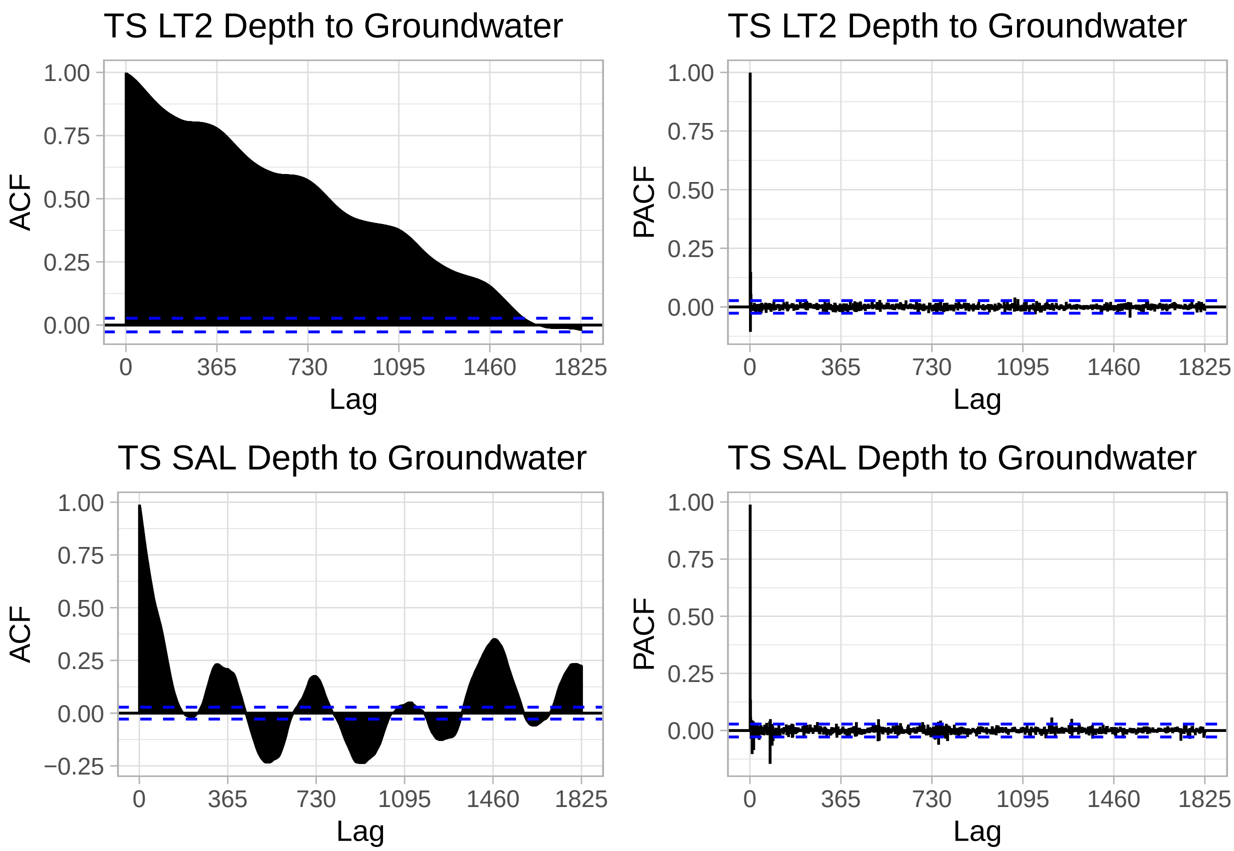


Figure 4: ACF and PACFs for Depth to Groundwater at Wells LT2 and SAL

The decomposition of each well's depth to groundwater data is shown in Figures 5 and 6 below. This decomposition displays overall trend, seasonality, and the random components of the depth to groundwater data over time.

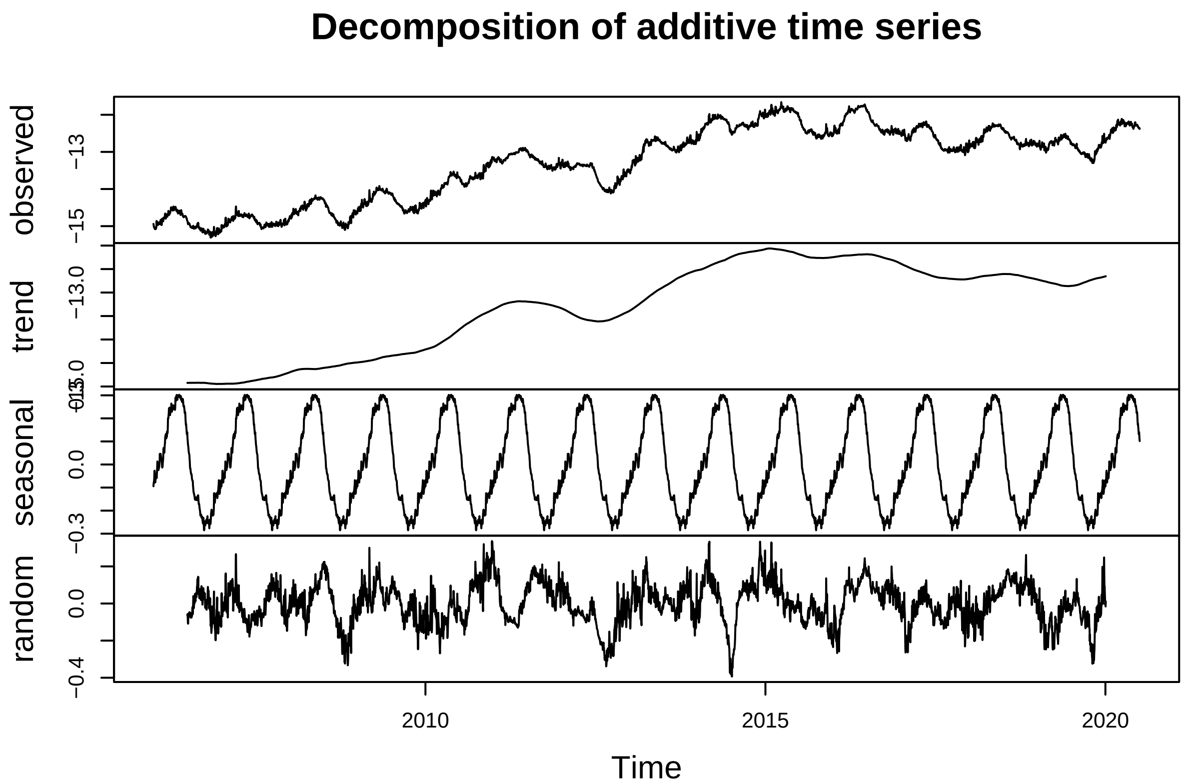


Figure 5: Decomposition of Depth to Groundwater at Well LT2

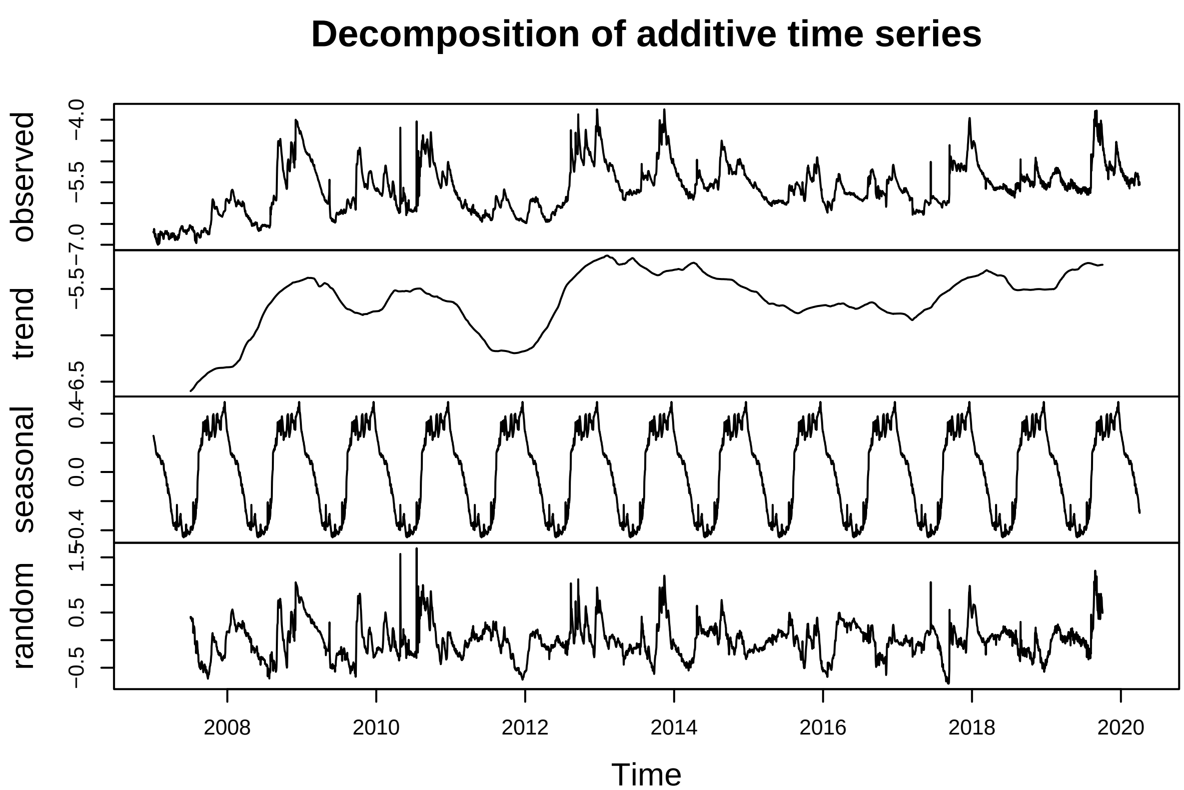


Figure 6: Decomposition of Depth to Groundwater at Well SAL

According to the decomposition, both wells show upwards trending depth to groundwater values over time. The ADF and Mann-Kendall tests were run on the de-seasoned timeseries to determine the nature of these trends. The LT2 well in the confined portion of the aquifer had a stochastic trend, while the SAL well in the unconfined aquifer had a deterministic trend. A summary of the statistical tests used to determine this are seen in Table 4 below. These different trends are likely due to the fact that wells are monitoring different kinds of aquifers, confined and unconfined.

Table 4: Trend Conclusions from the Augmented Dickey Fuller and Mann Kendall Test Results

	SAL North Well (confined)	LT2 South Well (unconfined)
ADF Test	p-value = 0.01	p-value = 0.9466
Result	Reject Null	Fail to Reject Null
MK Test	p-value =< 2.22e-16	NA
Result	Reject Null	NA
Conclusion	Deterministic Trend	Stochastic Trend

Models were then fit to the daily and monthly depth to groundwater time series data for each well while holding out a year of data. Model performance was then evaluated on the final year of held out data.

The results of the Auto Sarima (SARIMA) (using monthly data), the Exponential Smoothing State Space (ETS) (daily data), the State Space Exponential Smoothing (SSES) (daily data), and the Neural Network (NN) (daily data) models compared to the observed data are shown in Figures 7 and 8 below. A summary of the model performance metrics against the held-out test data are in Tables 5 and 6.

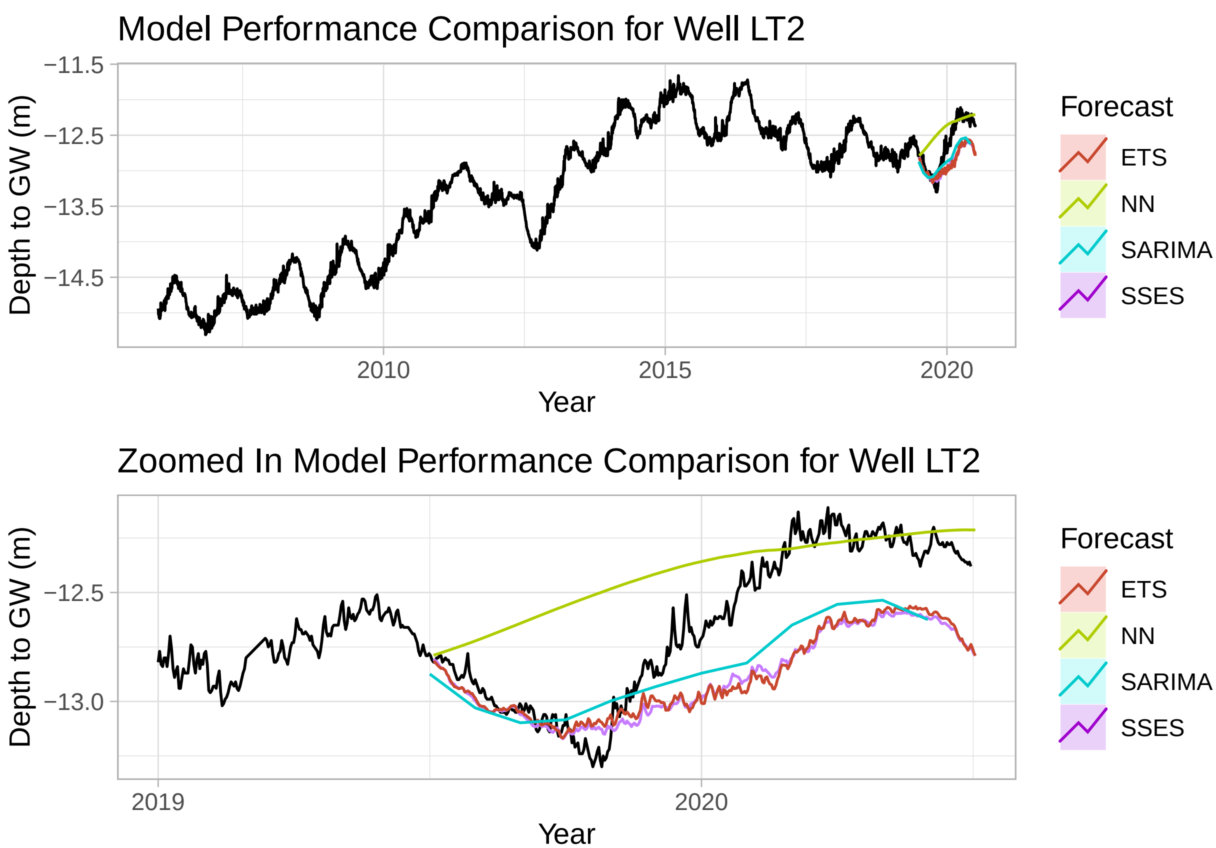


Figure 7: Model Performance Comparisons for Well LT2

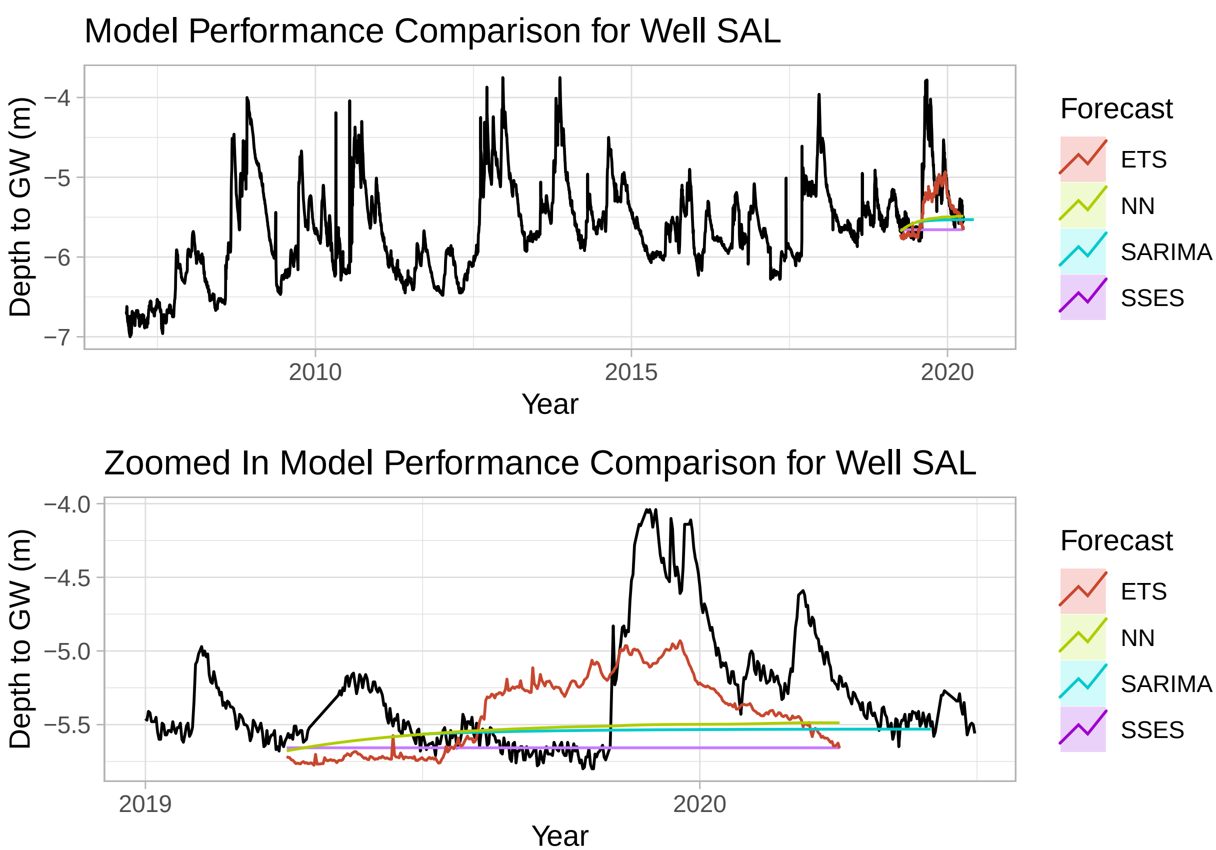


Figure 8: Model Performance Comparisons for Well SAL

Table 5: Forecast Accuracy for Seasonal Data at Well LT2

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
SARIMA	0.18533	0.25429	0.20708	-1.50064	1.66526	0.78460	1.76030
ETS	0.22956	0.30488	0.25138	-1.85643	2.02133	0.97974	8.71599
SSES	0.23689	0.30485	0.25382	-1.91339	2.04135	0.97877	8.72128
NN	-0.23675	0.33732	0.25579	1.82497	1.98123	0.98750	9.15380

Table 6: Forecast Accuracy for Seasonal Data at Well SAL

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
SARIMA	0.28294	0.51523	0.35878	-6.16268	7.50082	0.47825	1.27277
ETS	0.18165	0.40155	0.24084	-4.12608	5.22038	0.96610	5.06406
SSES	0.41811	0.63229	0.43603	-8.99277	9.30487	0.98020	7.73355
NN	0.29771	0.54664	0.35591	-6.64153	7.66261	0.97911	6.77645

The SARIMA model resulted in the lowest root mean square error (RMSE) for the LT2 well, and the ETS model resulted in the lowest RMSE for the SAL well.

After these initial results, exogenous variables were incorporated into the SARIMA model to see if the performance on the hold-out data could be improved. Both rainfall and temperature were added as exogenous variables, and they both improved the model performance on the held-out data.

The SARIMA model results with exogenous variables are seen in Figure 9 below. Tables 7 and 8 compare performance metrics of the SARIMA model without exogenous variables with the models that included rainfall and temperature. At each of the wells, the SARIMA model with temperature as an exogenous variable had the lowest RMSE and mean absolute percentage error (MAPE).

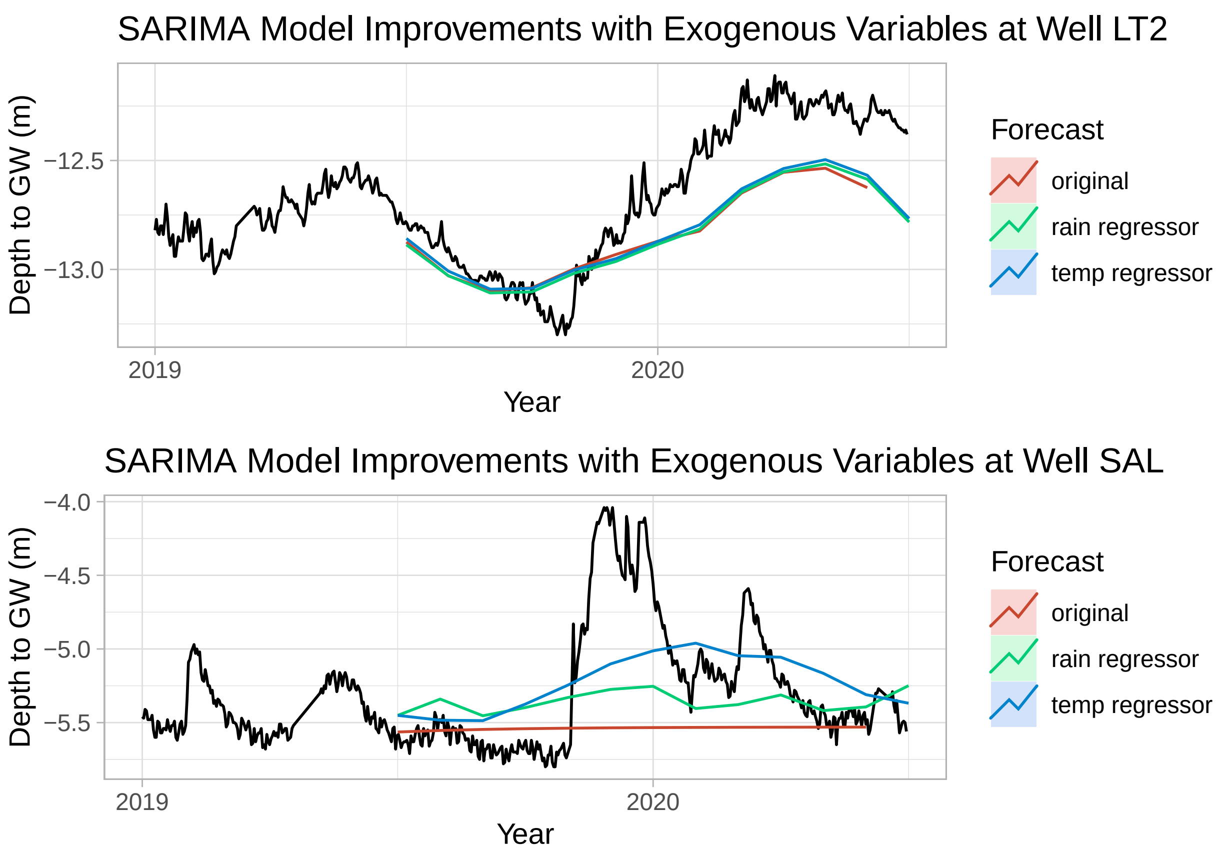


Figure 9: SARIMA Model Improvements when Exogenous Variables are Included at Wells LT2 and SAL

Table 7: Forecast Accuracy for Sarima with Regressors at Well LT2

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
SARIMA	0.18533	0.25429	0.20708	-1.50064	1.66526	0.78460	1.76030
SARIMA w/ RAIN	0.18855	0.25089	0.20714	-1.52366	1.66434	0.78531	1.73493
SARIMA w/ TEMP	0.17090	0.23875	0.19216	-1.38419	1.54510	0.78406	1.65290

Table 8: Forecast Accuracy for Sarima with Regressors at Well SAL

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
SARIMA	0.28294	0.51523	0.35878	-6.16268	7.50082	0.47825	1.27277
SARIMA w/ RAIN	0.11106	0.41086	0.32007	-2.80383	6.51817	0.42753	0.99058
SARIMA w/ TEMP	-0.03167	0.35584	0.28774	0.03923	5.75457	0.32779	0.84574

Summary and Conclusions

In summary, four different models were evaluated to predict depth to groundwater at two wells within confined and unconfined portions of the Auser Aquifer in Italy. These included a seasonal arima, exponential smoothing, state space exponential smoothing, and a neural network model. When exploring model performance without exogenous variables, ETS best predicted depth to groundwater in the confined aquifer, and the seasonal auto arima best predicted depth to groundwater in the unconfined aquifer, as determined by RMSE and MAPE scores on a final year of holdout data. The model performances are seen in Figures 7 and 8, and forecast accuracy metrics are displayed in Tables 5 and 6.

Incorporating exogenous variables into the seasonal auto arima model improved model performance for both wells, as seen in Figure 9 and the forecast accuracy metrics of Tables 7 and 8. Using temperature as an exogenous variable improved model accuracy the greatest for both of the wells analyzed in this project.

The well within the unconfined aquifer (SAL) had more significant improvement in model forecast when including exogenous variables than the well within the confined aquifer (LT2). This is likely due to the fact that groundwater levels within the unconfined aquifer fluctuate with greater frequency due to climatic factors. Rain can recharge unconfined aquifers faster than confined aquifers, and extreme heat would cause more evaporation in an unconfined aquifer than a confined one.

That being said, these climatic factors definitely influence depth to groundwater, so it makes sense that variables of temperature and rainfall both improved model performance when used as exogenous variables. There is room for further exploration of how these variables influence depth to groundwater predictions, as it is possible that there is a lag in the correlation between groundwater depth and rainfall that was unaccounted for in these models. This could be explored further in future work to improve the model even more.

Additionally, these results show that depth to groundwater is increasing with time, either stochastically or deterministically, in different portions of the aquifer, as seen in Figures 5 and 6 and Table 4. This means overall, groundwater levels are decreasing in this region. Combining these forecast models with future climate scenarios for temperature and rainfall to predict future groundwater depletion could help inform priority areas for groundwater conservation for utility companies.

In conclusion, it is valuable to understand the role temperature and rainfall data play in improving forecast model accuracy. These findings can help utilities prioritize the data that they are collecting at their sites to more accurately predict future aquifer storage.

Bibliography

Un world water development report 2022 'Groundwater: Making the invisible visible.' (n.d.). UN-Water. Retrieved April 18, 2024, from https://www.unwater.org/news/un-world-water-development-report-2022-

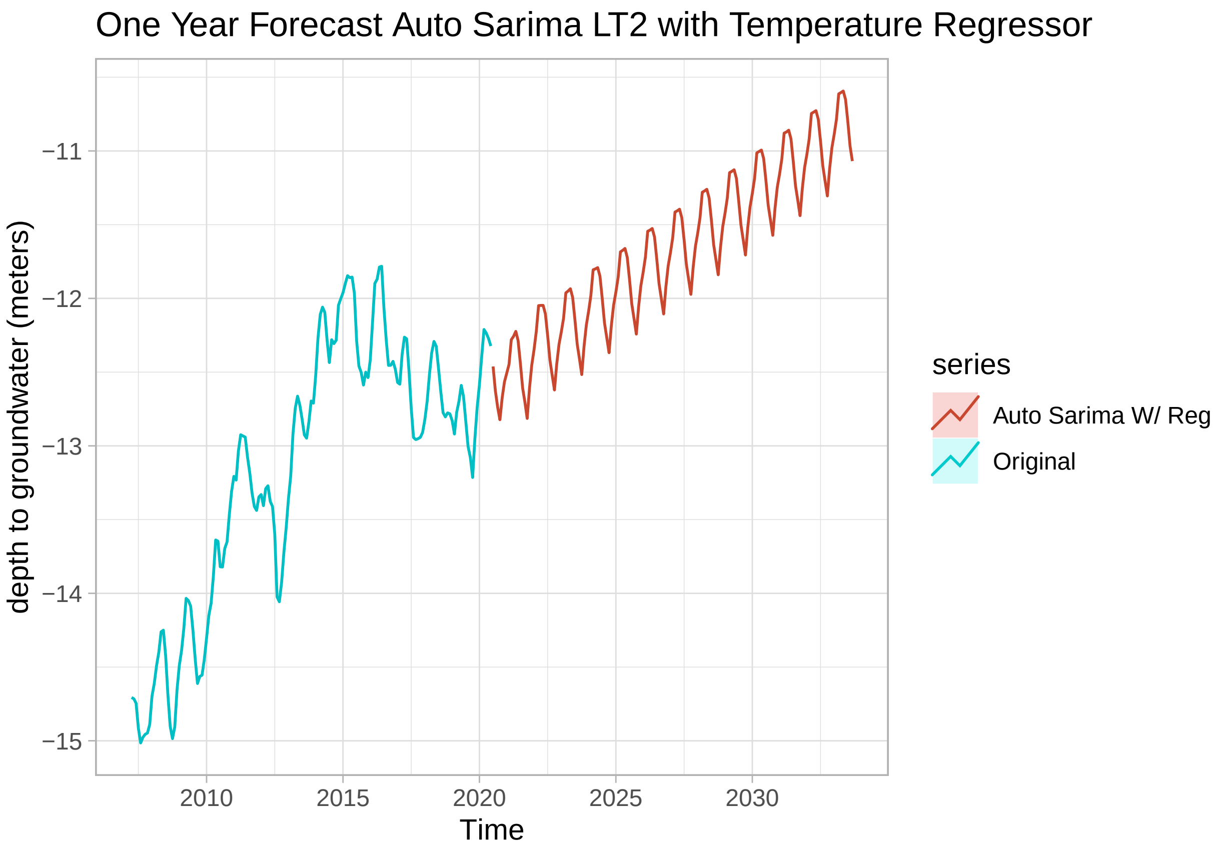
%E2%80%98groundwater-making-invisible-visible%E2%80%99

Antimo Musone, Aredhel Bergström, Federico, Luisa Marotta, Maggie, Maurizio Lucchesi. (2020). Acea Smart Water Analytics. Kaggle. https://kaggle.com/competitions/acea-water-prediction

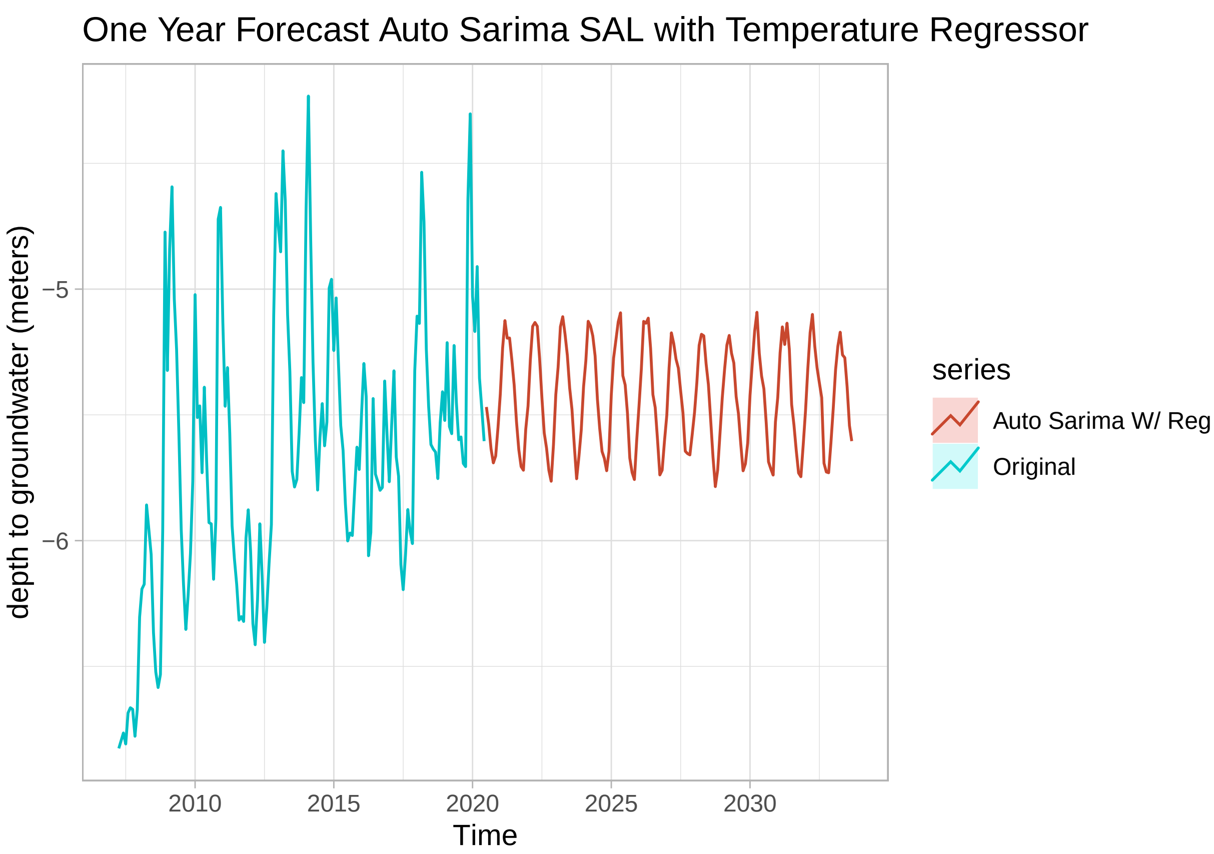
Appendix

For further analysis, we attempted to run a forecast of the seasonal arima one year into the future to see what future groundwater availability in these two wells would look like. We ran into an issue where our forecast with temperature as an exogenous variable ended up forecasting 12 years into the future instead of 12 months. After multiple attempts, we were not able to figure out how to rectify the issue. Although these forecasts are not what we had hoped for, we still found the results interesting and wanted to include them in our report. Below we have shown our code and the resulting plots for reference.

```r
ts_LT2_monthly_reg <- tsclean(ts(LT2_for_reg[,2],start=c(2007,04,01),frequency=12))
ts_SAL_monthly_reg <- tsclean(ts(SAL_for_reg[,2],start=c(2007,04,01),frequency=12))
ts_rain_monthly_reg <- tsclean(ts(rain_reg_monthly[,2],start=c(2007,04,01),frequency=12))
ts_temp_monthly_reg <- tsclean(ts(temp_reg_monthly[,2],start=c(2007,04,01),frequency=12))

auto_LT2_future <- auto.arima(ts_LT2_monthly_reg, xreg=ts_temp_monthly_reg)

auto_LT2_future_forecast <- forecast(auto_LT2_future, h=12, xreg=ts_temp_monthly_reg)

#plot model + observed data
autoplot(ts_LT2_monthly_reg, series = "Original") +
  autolayer(auto_LT2_future_forecast, series = "Auto Sarima W/ Reg", PI = FALSE) +
  ylab("depth to groundwater (meters)") +
  ggtitle("One Year Forecast Auto Sarima LT2 with Temperature Regressor") +
  theme_light()
```




```r
auto_SAL_future <- auto.arima(ts_SAL_monthly_reg, xreg=ts_temp_monthly_reg)

auto_SAL_future_forecast <- forecast(auto_SAL_future, h=12, xreg=ts_temp_monthly_reg)

#plot model + observed data
autoplot(ts_SAL_monthly_reg, series = "Original") +
  autolayer(auto_SAL_future_forecast, series = "Auto Sarima W/ Reg", PI = FALSE) +
  ylab("depth to groundwater (meters)") +
  ggtitle("One Year Forecast Auto Sarima SAL with Temperature Regressor") +
  theme_light()
```


One Year Forecast Auto Sarima SAL with Temperature Regressor
depth to groundwater (meters)
−5
−6
2010
2015
2020
2025
2030
Time
series
Auto Sarima W/ Reg
Original