

Article

Not peer-reviewed version

Guideline for Developing Drinking Water Capacity Analysis Report Using Time Series Linear Model to Predict Future Demand

[Dessy Owiti](#)*, Hulya Julie Yazici, Justin Mahon

Posted Date: 27 May 2026

doi: 10.20944/preprints202605.1873.v1

Keywords: capacity analysis report; time series linear model; drinking water; public water systems



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Guideline for Developing Drinking Water Capacity Analysis Report Using Time Series Linear Model to Predict Future Demand

Dessy Owiti ^{1,2,*}, Hulya Julie Yazici ² and Justin Mahon ¹

¹ City of Fort Myers, Utilities Department, Florida, USA

² Florida Gulf Coast University, Florida, USA

* Correspondence: DOwiti@fortmyers.gov or daowiti4139@eagle.fgcu.edu; Tel.: +1-239-321-7224

Abstract

A capacity analysis report (CAR) is an essential document that assists public water system owners in prioritizing water infrastructure and resources for future capital improvements. The United States Environmental Protection Agency states that public water systems should develop a CAR when average daily production approaches or exceeds a certain percentage of the system design capacity. Different states have different CAR requirements; this report is a guide for developing a CAR for drinking water facilities based on State of Florida requirements. This document will also demonstrate how to use a water demand time series predictive model to forecast future demand. An illustration of how to develop two models is provided. Both models use time as the predictor/independent variable and water demand/consumption as the dependent variable. The first model uses full data to determine the linear regression equation, while the second model uses partitioned data to develop the linear regression equation. The model with the lowest Mean Absolute Error (MAE) is selected for future demand predictions. City of Fort Myers Water Treatment Plant data was used to demonstrate how to develop a CAR and demand forecasting models.

Keywords: capacity analysis report; time series linear model; drinking water; public water systems

1. Introduction

Assessing current and future water consumption is a critical element of managing water supply infrastructure and water resources [1]. A capacity analysis report (CAR) is an essential document that assists public water system owners in prioritizing water infrastructure and resources for future capital improvements. Public water systems (PWSs) develop CARs to facilitate planning, design, and expansion efforts to ensure that current and future demands (also referred to as consumption) are met [2]. Water consumption evaluation can be used to develop revenue projections to be incorporated into municipal budgets and justify major investments in water infrastructure [3].

Content requirements of a CAR vary by state. For instance, Washington state Department of Health [4] and Delaware Health and Social Services [5] require a CAR to include source and treatment capacity, managerial capacity and financial capacity. This document will illustrate how to develop a CAR for drinking water facilities based on State of Florida requirements outlined in rule 62-555.348 Florida Administrative Code (F.A.C.). Public Water systems in other states may use this document as a guide for developing a CAR, and, if required, refer to the United States Environmental Protection Agency (EPA) Sanitary Survey Learner's Guide section 14 to develop the managerial and financial capacity sections [6].

The EPA Guide states that PWSs should have a plan for capital improvements once the average daily production approaches or exceeds a certain percentage of the design capacity of major PWS components, such as source, treatment, or storage facilities [6]. In the State of Florida, a CAR is required when the water produced at the water treatment plant exceeds 75 percent of the permitted

design capacity listed in the facility's construction permit. Rule 62-555.348 F.A.C. states that utilities that serve 350 or more persons or 150 or more service connections must submit a CAR when the total maximum-day quantity of finished water produced by all treatment plants connected to a PWS, including water produced to meet any fire-flow demand, exceeds 75 percent of the total permitted maximum-day operating capacity (permitted design capacity) of the PWS [2]. The rule indicates that unusually extreme values should be excluded. Water produced to meet any demand that the supplier of water documents to be highly unusual and nonrecurring should be excluded from these calculations [2].

The Florida Department of Environmental Protection (FDEP), the regulatory agency overseeing potable water and wastewater, provides step-by-step instruction for developing a CAR for wastewater but not drinking water [7]. The general chapters outlined in the wastewater CAR are transferable to a potable water CAR. However, the intricate details required within each chapter vary drastically, this report will bridge the gap. In addition, the wastewater CAR proposes using population and gallons per capita per day to project future flows.

The objective of this study is to explain the requirements of a CAR outlined in rule 62-555.348 F.A.C. in straightforward terms and describe how to forecast demand using linear time series predictive analytics. This document will demonstrate how to use a water demand time series predictive model to forecast future demand. In this report, the graphical and tabular examples are based on data from the City of Fort Myers PWS. In addition, data used to illustrate water demand predictive models are historical data from the City of Fort Myers PWS. The data is public record that can be retrieved from the FDEP Information Portal website [8].

2. Florida Capacity Analysis Report Required Components

The City of Fort Myers (City) is in Southwest Florida along the Caloosahatchee River and is neighbored by the City of Cape Coral to the west, Lehigh Acres to the east, and unincorporated Lee County to the south [9]. See the City Boundary in Figure 1. The City owns and operates one water treatment facility, the City of Fort Myers PWS (PWS number 5360102) [10]. The water treatment facility is located at 2751 Jacksonville Street, Fort Myers, Florida 33916, in Lee County, Florida [10]. The location of the facility is shown in Figure 2, the facility receives water supply from the Upper Floridan Aquifer, which is treated using reverse osmosis (RO) treatment [10].

In Florida, the required content of a drinking water CAR can be summarized as shown in the list below. Managerial capacity and financial capacity are evaluated during the permitting process of a PWS construction application and in the subsequent sanitary survey inspections post construction, and thus not required in a CAR report.

- Evaluation of the capacity of all sources, treatment, and storage facilities connected to the PWS.
- Comparison between the maximum-day water demand (including fire-flow demand if fire protection is being provided) versus permitted design capacity of the water treatment plant(s).
- Assessment of the finished-water storage needed (including fire storage if fire protection is provided) versus the existing total useful finished-water storage capacity.
- Lastly, expansion or new construction recommendations.

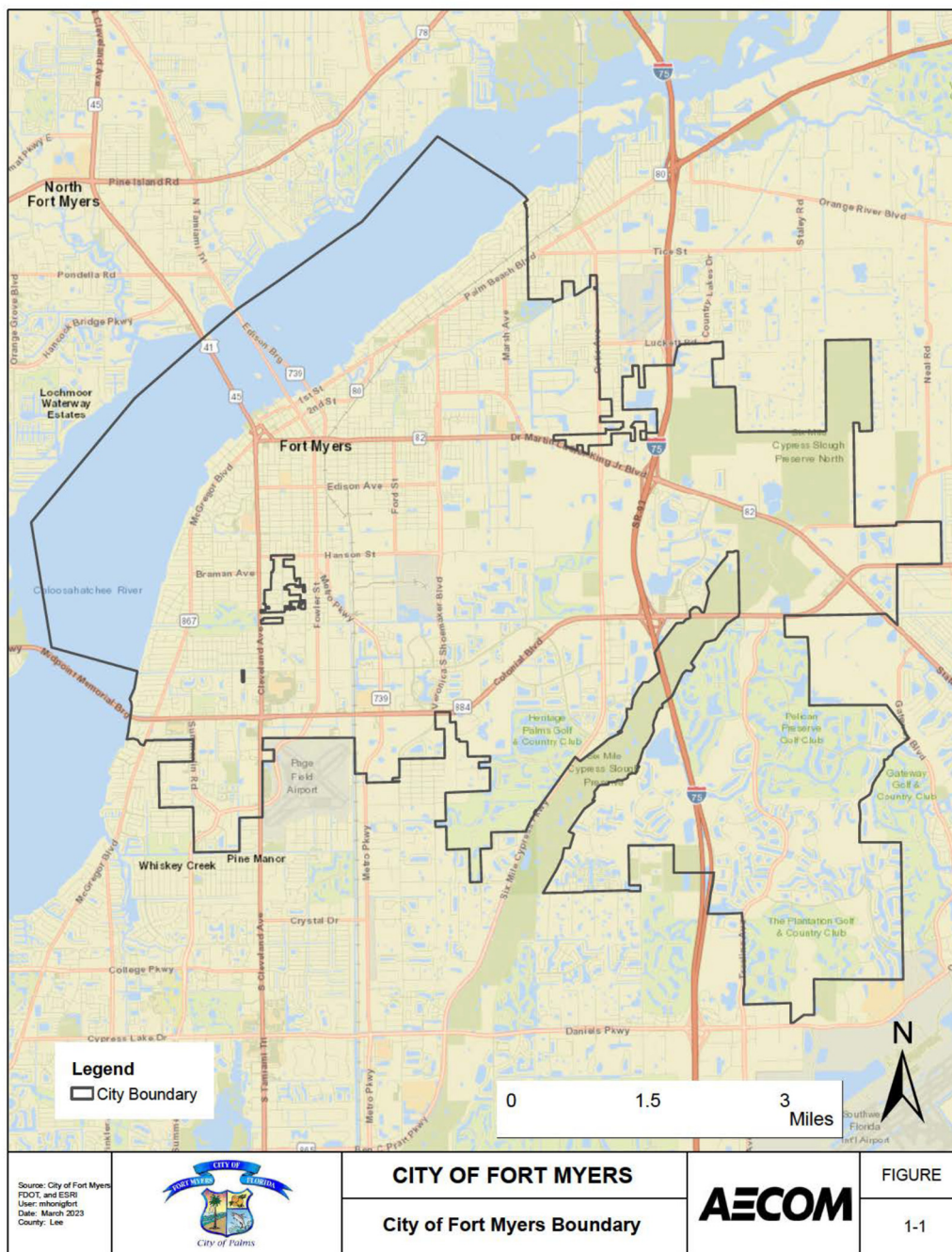


Figure 1. The City of Fort Myers boundaries [9].



Figure 2. Location of City of Fort Myers water treatment plant [9].

2.1. Capacity Analysis Report Submittal Timelines

The Florida rule 62-555.348(3)(a) and (b) F.A.C. [2] specify when an initial report is required and when the subsequent updated reports are required. The CAR submittal timelines requirements are described below.

- An initial CAR is required six months after the month in which the total maximum-day quantity of finished water produced (MDP) by the treatment plant(s) first exceeds 75 percent of the permitted design capacity.
- An updated CAR is not required if the initial or updated report indicates that at build-out, maximum-day water demand, including fire-flow demand (MDD + FF) will not exceed permitted design capacity and/or finished-water storage, including fire storage (Storage + FF) will not exceed the existing total useful finished-water storage capacity (useful storage capacity).
- An updated CAR is required within 5 years after the submittal of the initial report if an initial or updated CAR indicates that MDD + FF will exceed permitted design capacity after 10 years and/or Storage + FF will exceed the existing useful storage capacity after 10 years.
- An updated CAR is required within 2 years after the submittal of the initial report if an initial or updated report indicates that MDD + FF will exceed permitted design capacity in 5 to 10 years and/or Storage + FF will exceed the existing useful storage capacity in 5 to 10 years.
- An updated CAR is required 1 year after the submittal of the initial CAR if an initial or updated report indicates that MDD + FF will exceed permitted design capacity in less than 5 years and/or Storage + FF will exceed the existing useful storage capacity in less than 5 years.

2.2. Existing Facility Description and CAR Certification Requirements

Rule 62-555.348(5) F.A.C. states that the CAR must be prepared by a person under the responsible charge of one or more professional engineers licensed in Florida and shall be signed, sealed, and dated by the professional engineer(s) in responsible charge [2].

Rules 62-555.348(4)(a), 62-555.348(4)(b) F.A.C. and 62-555.348(5) F.A.C. indicate that a CAR shall contain the information for the existing PWS.

- The capacity of each water treatment plant's source water facilities and treatment facilities.
- The permitted maximum-day operating capacity (permitted design capacity) and if applicable, permitted peak operating capacity of each plant.
- The useful capacity of each finished-water storage facility.
- The MDP and annual average-day quantity of finished water produced (AADP) by each plant during each of the past ten years or during each of the years the plant has been in operation, whichever is less.

2.2.1. Description of Treatment Process

The CAR should include a brief overview of the treatment process from source water (surface water or groundwater/well) to the finished water before distribution. A flow diagram showing the major treatment process units is preferred. If the CAR is an updated report, the report should include the submittal date of the last CAR [7] and any substantial changes. Figure 3 shows a flow diagram of the City of Fort Myers water treatment plant retrieved from FDEP information portal [8].

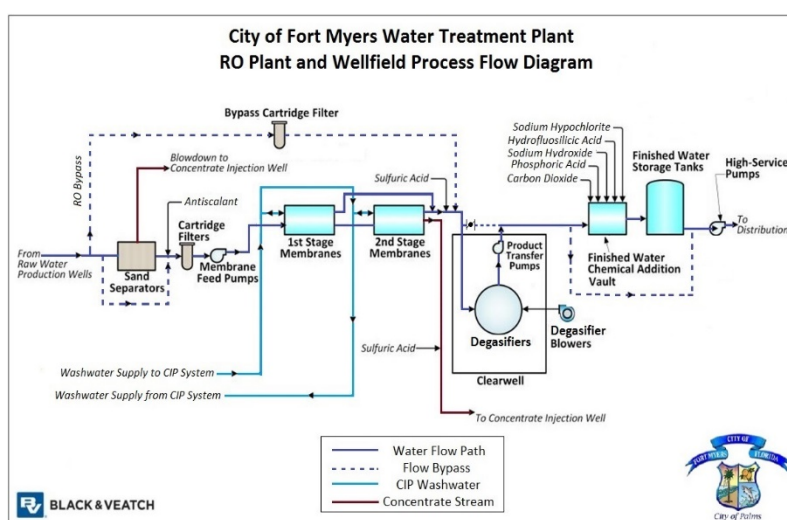


Figure 3. Flow diagram of treatment process at City of Fort Myers water treatment plant [10].

2.2.2. Description of Permitted Maximum-Day Operating Capacity (Permitted Design Capacity), Useful Capacity of Finished-Water Storage Facilities, Source Water Capacity and Capacity for Units Used in Treatment Process

The CAR should indicate the permitted design capacity listed in the most recent PWS construction permit. The permitted design capacity is based on firm capacity, which is the capacity of the entire PWS that is equivalent to the MDD + FF with the most restrictive treatment processing unit or largest equipment out of service [2].

The useful capacity of each finished-water storage facility must be included in the CAR. Useful capacity of finished-water storage facility is the amount of water available for use in a storage facility such as a ground storage tank without causing pump suction [10]. Table 1 illustrates how to list finished water storage tanks capacity. Rule 62-555.320 (19)(a) F.A.C., stipulates that the total useful finished-water storage capacity (excluding storage capacity for fire protection) must equal 25 percent (25%) of the system's MDD without fire-flow [2].

Table 1. City of Fort Myers water treatment plant finished water storage tanks [11], volume in Million Gallons (MG).

Parameter	Value
Type	Above-Ground Storage
Quantity	3
Diameter	160'
Wall Height	33'-3"
Volume (each)	5 MG
Total Volume	15 MG
Useful Capacity (each)	4.2 MG
Total Useful Capacity	12.6 MG

When listing capacity for each PWS component, it is important to use the actual physical capacity rather than the manufacturer's original rating, as equipment performance may decline over time due to deterioration. Each source of water capacity (surface water or groundwater/well) must be listed in the CAR. Table 2 shows an example of how to illustrate source water capacity. A similar table should be developed to show the capacity of all pumps used by PWS. On the source water table, if groundwater is the source, the capacity with the largest well-out of service must also be listed. Rule 62-555.315(3) F.A.C. states that if the water system is a community system serving 350 or more persons or 150 or more service connections, the total well capacity with the largest producing well out of operation shall equal at least the design average daily water demand (ADD), and preferably the MDD.

Table 2. City of Fort Myers source water capacity [11], in Gallons per Minute (GPM) and Million Gallons per Day (MGD).

Well Name	Well Diameter (inches)	Total Depth (feet)	Cased Depth (feet)	Pump Capacity (gpm)	Well Capacity (gpm)	Year Drilled	Use Status
P-1	16	775	464	1,750	935	1998	In Service
P-2	16	775	432	1,750	1,035	1999	In Service
P-3	16	831	510	1,750	300	2000	In Service
P-4	16	825	510	1,750	500	2000	In Service
P-5	16	832	480	1,750	339	2000	In Service
P-6	16	805	445	1,750	295	2000	In Service
P-7	16	800	445	1,750	417	2000	In Service
P-8	16	775	465	0	0	2004	Out of Service
P-9	16	750	470	0	0	2006	Out of Service
P-10	16	840	450	0	0	2006	Out of Service

P-11	16	840	465	1,750	215	2006	In Service
P-12	16	710	465	1,750	1,330*	2006	In Service
P-13	16	700	575	1,000	130	2006	In Service
P-14	16	680	440	1,000	570	2006	In Service
P-15	16	680	490	1,000	470	2007	In Service
P-16	16	720	460	1,000	160	2007	In Service
P-17	16	720	460	1,000	658	2006	In Service
P-18	16	685	485	0	0	2008	Abandoned
P-19	14	693	493	500	220	2016	In Service
P-20	14	720	492	500	468	2016	In Service
P-21	16	720	505	500	500	2022	In Service
P-22	16	716	505	500	500	2022	In Service
P-23	16	754	515	500	500	2024	In Service
P-26	16	645	505	500	500	2024	In Service
P-27	16	800	510	500	500	2023	In Service
P-28	16	673	510	500	500	2023	In Service
P-1S	16	700	465	600	600	2023	In Service
P-2S	16	720	470	500	500	2022	In Service
P-3S	16	575	450	500	500	2024	In Service
P-4S	16	520	430	500	500	2024	In Service
P-5S	16	720	510	500	500	2022	In Service
Total Well Capacity (gpm)					13,642	*Capacity with Largest well out of service = 1330gpm/17.73GPD	
Total Well Capacity (MGD)					19.64		

The capacity for units used in treatment process and all pumps should be listed. Table 3 provides an example of how to illustrate source water capacity. Table 2 shows an example of how to describe each unit capacity.

Table 3. City of Fort Myers unit process capacity [10], in Million Gallons per Day (MGD).

Process Unit	# of Units	Description	Unit Capacity (mgd)	Total Capacity (MGD)
Sand Separators	2	Hydac International; conical slotted tubes; 50 micron	15.52	31.04
Cartridge Filters	4	Fulflo Filter; 296 polypropylene elements per vessel; 1 micron	7.24	28.96

Membrane Feed Pumps	4	Goulds; horizontal, split case centrifugal; 500' TDH; 500 hp	4.32	17.28
RO Membrane System	8 - 1 st Stage 4 - 2 nd Stage	Train = 2 - 1 st stage & 1 - 2 nd stage; 8.5" vessels (3 trains); 8" vessels (1 train); 70% recovery Bypass = 0.16 MGD	2.72 + Bypass	11.04
Veolia Proflex 108 RO Membrane Units ¹	2 - 2-Stage Skids	8-inch vessels 80% Recovery	0.5	1.0
Degasifiers ²	3	2 legacy, 1 new degasifier; Jaeger 3.5-inch Tri-Pack packaging; 3 centrifugal blowers; 60-000 cfm	6.67	20.01
Product Transfer Pumps	4	Vertical Turbine; 150' TDH; 125 hp	6.77	27.08

Presenting the capacity for useful finished water storage, source water, and treatment process units helps clearly identify the bottleneck to confirm the permitted design capacity and to assess whether the most limiting factor can meet future demand.

2.2.3. Max-Day and Annual Average-Day Produced

In this section, a table or a graph should be developed to show the MDP and annual average-day quantity of finished water produced (AADP) by each plant for the past 10 years or during each of the years the plant has been in operation (whichever is less). A record of water produced by the PWS can be found in the Monthly Operating Reports (MOR) submitted to the regulatory agency. In Florida, these records can be retrieved from FDEP information portal [8]. A historical record of the finished water produced provides insight into the treatment plant (s) capabilities.

2.3. Water Demand Projections and Plan of Action Required

The Florida Rule 62-555.348(4) parts (c) through (f) F.A.C. and 62-555.348(6) F.A.C., specify information required for future water demand projections and plan of action based on those projections [2]. The requirements are listed below.

- Projected total water demands that show total annual average daily demand and total maximum-day demand (including fire-flow demand if fire protection is being provided) – for at least the next ten years and projected total finished-water storage needed (including fire storage if fire protection is being provided) for at least the next ten years.
- An estimate of the time required for maximum-day water demand (including fire-flow demand if fire protection is being provided) to exceed the total permitted maximum-day operating capacity of the plant(s) and an estimate of the time required for finished-water storage need (including fire storage if fire protection is being provided) to exceed the existing total useful finished-water storage capacity.
- Recommendations for a new or expanded source, treatment, or storage facilities.
- A recommended schedule showing dates for design, permitting, and construction of recommended new or expanded source, treatment, or storage facilities.

Rule 62-555.348 (6) F.A.C. states that if an initial or updated source/treatment/storage capacity analysis report indicates that MDD+FF will exceed the permitted design capacity of the water treatment plant(s) in less than 5 years or that finished-water storage need (including fire storage) will exceed the existing total useful finished-water storage capacity in less than 5 years, a document showing a schedule/ dates for design, permitting, and construction of recommended new or expanded source, treatment, or storage facilities shall be submitted with the report [2]. The

documentation must also include a written statement that is signed by an authorized representative of the supplier of water and that certifies the supplier is meeting, and intends to meet, the report's recommended schedule for design, permitting, and construction of recommended new or expanded source, treatment, or storage facilities [2].

2.3.1. Projected Total Water Demands and Future Needs for the Finished Water Storage

After demonstrating the capacity of existing PWS components, the next step is to project total water demands listed in Rule 62-555.348(4)(c) F.A.C. The rule requires a projection of total annual average daily demand (AADD) and MDD + FF for at least the next 10 years. In addition, the total finished-water storage needed for the next 10 years or more must also be defined. As stated earlier, the minimum requirement for finished water storage is that total useful finished-water storage capacity (excluding storage capacity for fire protection) must equal 25% of the system's MDD without fire-flow.

To satisfy these regulatory requirements, utilities must apply an appropriate predictive approach to estimate future water demand and storage needs based on historical data. Predictive analytics provides several modeling techniques that can be used to evaluate trends and support long term planning decisions. Selection of a suitable model depends on the type of data being analyzed and the desired outcome of the analysis.

There are five types of predictive models that are used in predictive analysis. They include: classification model whereby data is categorized into predefined groups based on historic data; clustering model which identifies similarities in data and creates groups based on these similarities; outliers model which identifies anomalies relative to historic data; forecast model whereby predictions are made about numeric data based on historic data; time series model which makes predictions about numeric data based on historic data with time as a factor [12].

This study demonstrates how a time series model can be used for predicting future water demand. Historically, City of Fort Myers drinking water CARs applied the indirect method of projecting future demand. This involved calculating per capita demand and forecasting population by multiplying the projected population with per capita demand to predict future consumption [9]. The time series method can be used as an alternative method for making predictions. It is a direct method of predicting future demand that illuminates the inherent autocorrelation structure of a water use pattern over time [13].

The time series model differs from population and per capita method in that it requires less effort to estimate demand specific to the PWS and its service area customers. When using population and per capita demand to make water consumption predictions, the population estimated within City of Fort Myers boundaries had to be adjusted to account for areas within the City of Fort Myers boundaries that are served by Lee County Utilities PWS [9]. When using time series model, the time (dates/years) column is the predictor variable while the demand column is the output variable, no additional adjustments are required on the output variable. In addition to its simplicity, the time series model captures factors that influence demand and that are specific to a PWS such as; operational characteristics unique to the water treatment plant, irregularities in water consumption, weather conditions (such as drought or rain), seasonal residents, seasonal variations, demographics, population, cultural and leisure events that have an impact on demand [3]. Time series analysis is one of the most widely used techniques for water demand forecasting [13]. Future water demand is determined based on historical time series patterns [1]. Toho Water Authority Poinciana Water Supply System, PWS ID 3494429 submitted a CAR to FDEP in December 2024 which used timeseries linear extrapolation as a forecasting method for drinking water [14].

3. Materials and Methods for Linear Time Series Predictive Model

The second objective of this study is to outline steps for developing linear time series model to predict water demand using excel [15] and Excel Analytical Solver Data Science software [16]. Two methods can be applied to determine the linear regression line equation to be used to make water

demand predictions. The first model uses the full 10-year historical data set to develop the linear regression line equation while the second model uses partitioned historical data to develop the linear regression line equation to be used for water demand projections. The Mean Absolute Error (MAE) from the two models should be compared, the model with the lowest MAE will be used to make predictions. Error term MAE is a statistical method used to determine how accurate predictions are. The term "Absolute" in this case means that when calculating averages, the positive or negative prefix of a numerical value are not included in the calculation. The MAE is calculated as the sum of (absolute actual/ true value – absolute predicted value) the historical demand data. The result provides a clear indication of how much, on average, the predictions deviate from the actual values [17]. Lower MAE values indicate better model performance, meaning the predictions are closer to the actual values; an MAE of zero means perfect predictions [17].

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

where:

- N is the number of observations.
- y_i represents the actual value for the i -th observation.
- \hat{y}_i represents the predicted value for the i -th observation.

Figure 4. Mean Absolute Error mathematical definition [17].

3.1. Data Cleaning

The first step is to identify outliers in the demand data and remove them prior to the time series predictive model development. A Z score provides a standardized way of assessing the average distance of a data point from the mean [18]. Positive Z-scores indicate values above the mean, while negative scores indicate values below the mean [18]. In this report the Z score was set at ± 2 . A Z-score of 0 indicates that each data point is exactly at the mean, a Z-score of ± 2 means each data point is 2 standard deviations away from the mean [18]. Data points with Z-score greater than positive 2 or less than negative 2 were considered outliers, meaning the data points are significantly different from the rest, thus indicate measurement errors or special occurrences such as significant water main breaks [18]. Instructions below describe how to identify outliers using excel [15].

- Create 5 columns. The first column (A) should be the date column, 2nd column (B) should be the daily demand column, 3rd column (C) should be the average demand column, 4th column (D) should be the standard deviation column and the 5th column (E) should be the Z score column.
- Under the average column (C) in cell C2 calculate the average of the whole demand data set using formula =AVERAGE (B first cell: B last cell).
- Under the standard deviation column (D) in cell D2 calculate the standard deviation of the whole demand data set using formula =STDEV (B first cell: B last cell).
- Under the Z score column (E) in cell E2 calculate the Z score of the first demand value/ data point in cell E2 using formula =STANDARDIZE(B2, \$C\$2,\$D\$2). Click the fill handle in cell E2 to apply the formula to the subsequent cells.
- To easily identify the outliers, use conditional formatting to identify values greater than 2 or less than -2. Select all values under the Z score column (E). select conditional formatting under the styles group. Select highlight cells rules. Select the greater than option. Enter the value 2 under format cells that are greater than. On the next tab select light red fill with dark red text from the drop-down menu. Click ok. Repeat the formatting step but the second time select less than

option and enter -2 under format cells tab then change color to green fill with dark green text. Review the flagged values to ensure that they are true outliers before manipulating the data points.

- Next step is to use the PivotTable to determine monthly averages as outlined in rule 62-555.348(4) F.A.C. Select the whole data set. Click insert tab, then the PivotTable drop down menu. Select the From Table/Range option. Select existing worksheet option. Select location as E2 and click ok. Move/ drag date to the Row section and Demand to the Values section on the right panel. Select the demand drop down menu and select value field settings. Under summarize value field by, select the average option. Unselect quarterly option located on the top panel of the pivot table fields section to remove it from the Rows section. Expand the Year in the pivot table to see the monthly demand values for each year. Create a new excel sheet with the monthly averages. If the average demand data includes fire flow be sure to exclude fire flow data.
- Repeat the steps above, select the demand drop down menu and select value field settings to get the Max values for each month. For the Max data be sure to include fire flow data.

3.2. Model 1: Using the Full Data to Develop the Linear Regression Line Equation to Be Used to Make Water Demand Predictions

Below is a description of the procedure for developing the linear time series regression line equation using full data (after data cleaning has been completed) with Data Analysis Tools in excel [15].

Select Data tab, then Data analysis tool tab and scroll down the pop-up box to select Regression and click ok. Select the input Y range tab and highlight the demand column including the column title. Select the input X range tab and highlight the date column including the column title. Select labels. Select confidence level and leave the default 95%. Do not select Constant is zero option. Select output range to have the results appear on the same excel sheet, select the tab and select cell M2 in the excel sheet. Results output will be entered starting from cell M2. Select all residuals then click ok tab.

In the results output, the R squared results indicate the goodness of fit; the closer the value is to 1 (or 100% if using percentages), the higher the explanation power of the independent variable on the dependent variable [19]. A high R squared indicates a good linear fit model. A low R squared may be an indication that a nonlinear model may be a better fit. The Intercept and X coefficient are listed under the column titled "coefficients". If this model is selected, the Y equation values would be retrieved from this column. Under the Residuals Output, the column titled "Residuals" represents the delta of actual demand values minus the predicted values of the historical data. Notice the predicted values are listed on the right side of the column titled "Observations".

The excel function =average(abs (First cell in residual column: last cell in the residual column)) is used to determine the Mean absolute Error (MAE) of the "Residuals" column. The linear time series analysis steps should also be repeated using maximum value data. For the Max data be sure to include fire flow data.

3.3. Model 2: Using Partitioned Data to Develop the Linear Regression line Equation to Be Used to Make Water Demand Predictions

Partitioning the data set into training and test datasets is sufficient [12]. However, in this report data was split into training, validation, and testing (also known as holdout) datasets. The dataset splitting mechanism can be created with specific criteria or by randomly selecting values and assigning them to either training, validation, or testing data [12]. Partitioning data allows for verification of accuracy of the model [12]. The training and validation dataset are used to build the predictive model, and their error terms are compared ensuring the model is not overfitted and that the identified trend reflects a true pattern [20]. The training dataset is employed to train the model to identify underlying patterns and relationships between the variables [20]. In this study, the validation

dataset is run against the developed model from the training dataset to evaluate its accuracy; if the values generated (predicted) closely match actual values, the model is considered valid [12]. If the model is valid, it is used to predict the dependent variable based on the independent variables [12]. The error term of the test dataset is used to assess future performance of the model [20].

The procedure for developing the linear regression line equation using the partitioned data is described below. Frontline Excel Analytical Solvers Data Science (ASDS) [16] is a tool that can be utilized for this process.

After data cleaning has been completed, copy the monthly dates into a new excel worksheet then duplicate the date column and paste it into a new column as numerical values rather than date data type. The ASDS software works better if all variables are in numerical data type. Click any cell with the data and select Data mining/Data science tab. Select the Partition tab under Data Mining group, (Careful NOT to select partition located under "Time series" group). Select standard partition. In the pop-up box select all variables under "Variables in input Data" by selecting the first variable, while holding the CTRL + shift keys, scroll down to select the last variable and then click the right arrow. Include all variables, even those that will not be used in developing the model). Select "specify percentages" Under "Partitioning percentages when picking up rows randomly". A larger dataset is used to train the model, therefore, allocate 70% of the data for training, 20% for validation and 10% for holdout (testing)). Do not change anything else in the pop-up box, click "ok". An excel sheet "STDpartition" will automatically open, scroll down to the "Partitioned Data" section.

The partitioned data is used to develop the model. Click any cell under the "Partitioned Data" section in the excel sheet "STDpartition" and select predict tab under data mining / data science select Linear Regression option. Next select the demand variable (target / output/ Y variable/ dependent variable) and move the demand to "Output Variable" by clicking the right arrow. Next select the date column with the numerical data type (predictor/X variables/ independent variable) and move it to "Selected variables". Click Next to get to the parameters tab. Select ANOVA, Analysis of Coefficients and Analysis of Residuals options. Click Next to get to "Scoring" tab, make sure to select "Summary Report" and "Detailed Report" under Training, Validation and Test/ hold data. You can skip "Simulation" tab and click finish. Multiple excel sheets will be generated.

As stated earlier, the R squared results indicate the linear fit of the model. A high R squared indicates that the model is a good linear fit model. A low R squared may be an indication that a nonlinear model may be a better fit. The Intercept and X coefficient are listed in the generated sheet titled "LinReg_Output". If this model is selected, the Y equation values would be retrieved from the coefficients column. The Residuals Output represents the delta of actual demand values minus the predicted values of the historical data. The residuals for the training, validation and test datasets can be found in excel worksheets titled "LinReg_TrainingScore", "LinReg_ValidationScore" and "LinReg_TestScore". Copy and paste the residual values into a new excel sheet and calculate the MAE as described in section 3.2 above. The linear time series analysis steps should also be repeated using maximum value data. For the Max data be sure to include fire flow data.

4. Results

4.1. Results for Model 1: Using the Full Data to Develop the Linear Regression Line Equation to Be Used to Make Water Demand Predictions

In full data analysis, the complete regression analysis results are as follows for City of Fort Myers PWS average demand data from 2015 to 2025, excluding fire flow.

- R squared = 82% an indication of a good linear fit.
- Intercept = -20.73 and the X coefficient was 0.000646. A line chart will render the same regression line equation.
- The error term MAE was 0.289 for this model.
- Y equation = $-20.737 + 0.000646(X)$.

In full data analysis, the complete regression analysis results is as follows for the City of Fort Myers PWS maximum demand data from 2015 to 2025, with fire flow.

- R squared = 79% an indication of a good linear fit.
- Intercept = -19.16 and the X coefficient was 0.000628. A line chart will render the same regression equation.
- The error term MAE was 0.297 for this model.
- Y equation = $-19.159 + 0.000628(X)$.

4.2. Results for Model 2: Using Partitioned Data to Develop the Linear Regression Line Equation to Be Used to Make Water Demand Predictions

In partitioned data analysis, the complete regression analysis results are as follows for City of Fort Myers PWS average demand data from 2015 to 2025, excluding fire flow.

- R squared = 80% an indication of a good linear fit.
- Intercept = -19.8 and the X coefficient was 0.000625.
- The error term MAE was 0.288 for this model.
- Y equation = $-19.799 + 0.000625(X)$.

In partitioned data analysis, the complete regression analysis results are as follows for the City of Fort Myers PWS maximum demand data from 2015 to 2025, with fire flow.

- R squared = 77% an indication of a good linear fit.
- Intercept = -18.33 and the X coefficient was 0.000609.
- The error term MAE was 0.301 for this model.
- Y equation = $-18.325 + 0.000609(X)$.

The model selected for future demand projection is based on the MAE error term. For the City of Fort Myers dataset, the results indicate that Model 1 (full data) average demand had an MAE value (0.289) that was very close to the Model 2 (partitioned data) average demand MAE value. In addition, model 1 max day demand MAE value (0.297) was lower than model 2 max day demand MAE value (0.301). Given the results, the model selected for future demand for the City of Fort Myers was the full dataset model.

Once the model is selected, demand prediction can be completed using the Y equation from the selected model. In the case of the City of Fort Myers, the Y equation for predicting future average demand (without fire flow) is $Y = -20.737 + 0.000646(X)$, and Y equation for predicting future Maximum demand (including fire flow) is $Y = -19.159 + 0.000628(X)$. In both cases Y represents the future demand while the X represents future monthly dates. After calculating the first predicted value, the results are converted into a numerical data type by selecting Number group drop down menu and choosing the Number option under category.

Rule 62-555.348(4)(c) F.A.C. requires a 10-year prediction. Once the predicted values are calculated graphs are developed to be incorporated in the CAR report. Figure 5 below shows an example of graphical representation of historical and predicted demands to be included in the CAR. The City of Fort Myers Water Treatment Plant received an approved permit dated October 24, 2024 that showed an increase in the permitted design capacity to 13.04 MGD following the installation of new RO units. This increase in permitted design capacity is illustrated in Figure 5 for the year 2024. In addition, the Water Treatment Plant is currently undergoing expansion, and upon the completion in 2028, the facility will be rerated to a capacity of 15 MGD as shown in Figure 5 below.

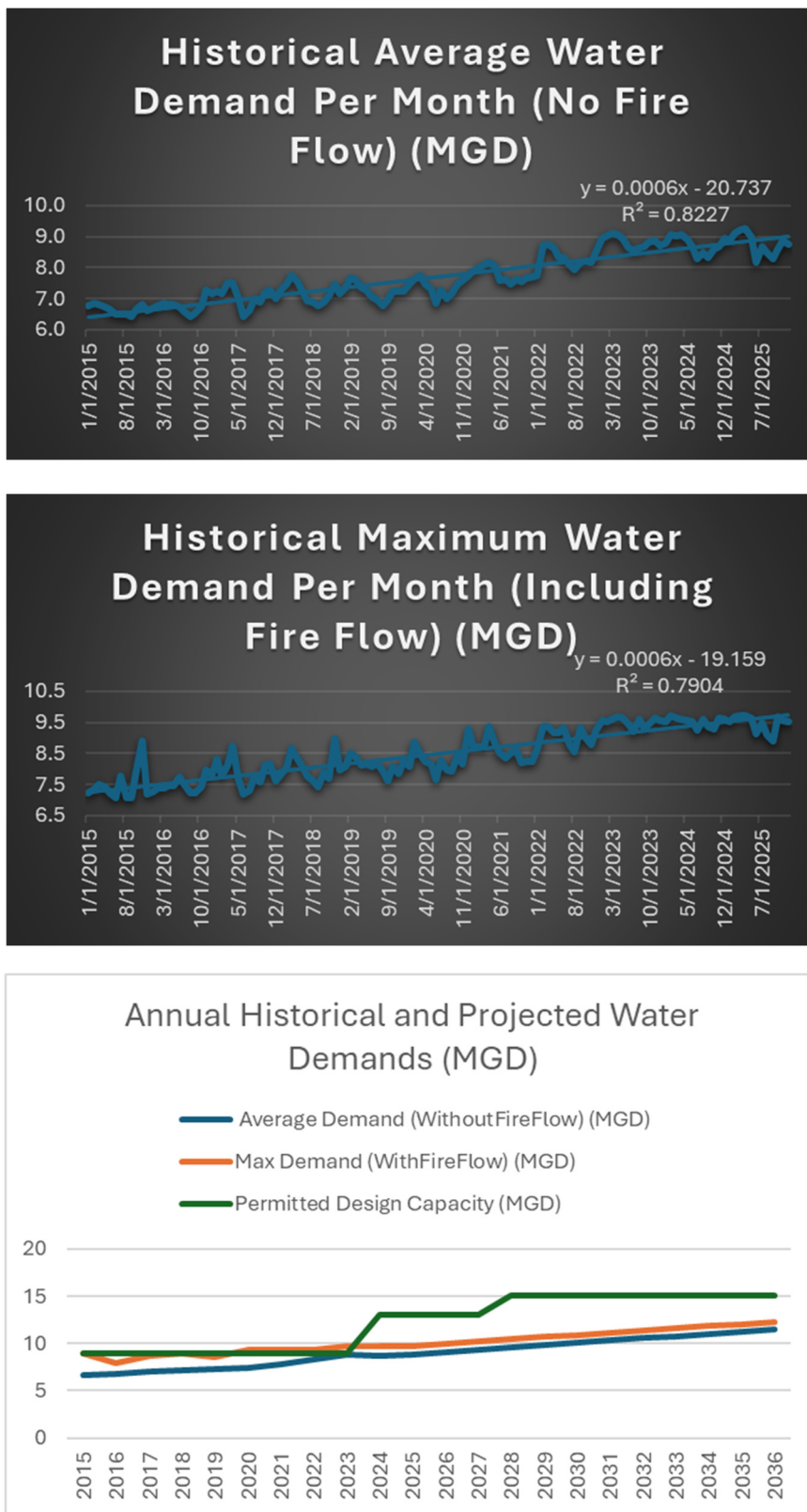


Figure 5. City of Fort Myers WTP Historical and predicted demands for the selected model for 2015 through 2036.

5. Discussion and Conclusions

This document provides a practical and comprehensive framework for developing a Capacity Analysis Report (CAR) for drinking water facilities based on State of Florida requirements outlined

in rule 62-555.348 F.A.C. Public Water systems in other states may use instructions in this document as a guide for developing a CAR and if required, use resources such as the EPA Sanitary Survey Learner's Guide to develop the managerial and financial capacity sections. The document also presents a direct method, linear time series predictive model, for forecasting future water demand. The method is well suited for PWSs because historical demand data is readily available, reducing the need for PWSs to outsource the development of a CAR. Future direction is to assess water demand projection using K Nearest Neighbors (KNN) Analytics Solver regression method.

Once demand projections have been established, the CAR should conclude by estimating when the MDD+FF will exceed the permitted design capacity and provide appropriate recommendations. The CAR could include a graph illustrating the estimated timeframe in which the MDD+FF is projected to exceed the existing permitted design capacity, as shown in Figure 5 above. Where permitted design capacity limitations are anticipated, recommendations for capacity rerating or expansion of PWS components should be provided. As started in this document, if the MDD+FF will exceed permitted design capacity in less than 5 years (or finished water storage will exceed the existing total useful finished-water storage capacity in less than 5 years); in addition to providing an implementation schedule, a certification from the authorized representative of the PWS will be required, stating that the supplier is meeting, and intends to meet, the report's recommended schedule for design, permitting, and construction of recommended new or expanded source, treatment, or storage facilities. If existing infrastructure and sources are determined to be sufficient for future demand, the CAR should provide a clear justification supporting that conclusion.

Through these evaluations and recommendations, the CAR serves as an essential planning and management tool for ensuring reliable, sustainable, and compliant drinking water service.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. The supporting information Excel file "Manuscript CAR 2026 Data analysis" submitted with the Manuscript.

Author Contributions: Conceptualization, D.O and J.M.; Formal analysis and methodology, D.O.; Validation, D.O. and H.Y.; Investigation, D.O.; Resources, D.O.; Data curation, D.O.; Writing—original draft preparation, D.O.; Writing—review and editing, H.Y. and J.M.; Visualization, D.O.; Supervision, H.Y and J.M.; Project administration, D.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. For access, contact the corresponding author.

Acknowledgments: The authors would like to thank Florida Gulf Coast University; Meagan Bledsoe, City of Fort Myers (City) Water Plant Supervisor; Jesse Victory, City Water Plant Superintendent; Jason Sciandra, Director of Utilities; Marty Lawing, City Manager; Mayor Kevin Anderson.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CAR	Capacity analysis report
F.A.C.	Florida Administrative Code
PWS	Public water systems
EPA	United States Environmental Protection Agency
FDEP	Florida Department of Environmental Protection
City	City of Fort Myers
MDP	Maximum-day quantity of finished water produced
AADP	Annual average-day quantity of finished water produced
MDD + FF	Maximum-day water demand, including fire-flow demand

AADD	Annual average daily demand
Storage + FF	Finished-water storage, including fire storage
MAE	Mean Absolute Error
MG	Million Gallons
MGD	Million Gallons per Day
GPM	Million Gallons per Minute

References

1. J. Stańczyk, J. Kajewska-Szkudlarek, P. Lipiński and P. Rychlikowski, "Improving Short-Term Water Demand Forecasting Using Evolutionary Algorithms," *Scientific Reports*, pp. <https://doi.org/10.1038/s41598-022-17177-0>, 2022.
2. Florida Department of State, *Florida Administrative Code & Florida Administrative Register Rule Chapter 62-555*, Tallahassee,, Florida,, 2014.
3. B. VanBerlo, M. Ross and D. Hsia, "Univariate Long-Term Municipal Water Demand," *arXiv (Cornell University)*, 2021.
4. Washington State Department of Health, "Water System Capacity Development," August 2013. [Online]. Available: <https://doh.wa.gov/sites/default/files/legacy/Documents/Pubs/331-283.pdf>. [Accessed 01 05 2026].
5. Delaware Health and Social Services, "Office of Drinking Water," 25 June 2021. [Online]. Available: <https://dhss.delaware.gov/wp-content/uploads/sites/10/dph/pdf/odwcapdevapdx.pdf>. [Accessed 01 05 2026].
6. United States Environmental Protection Agency, "How to Conduct a Sanitary Survey of Drinking Water Systems," August 2019. [Online]. Available: https://www.epa.gov/sites/default/files/2019-08/documents/sanitary_survey_learners_guide_508_8.27.19.pdf. [Accessed 01 05 2026].
7. Florida Department of Environmental Protection Domestic Wastewater Program, "Guidelines for Preparation of Capacity Analysis Reports," July 1992. [Online]. Available: <https://floridadep.gov/water/domestic-wastewater/documents/guidelines-preparation-capacity-analysis-reports>. [Accessed 01 05 2026].
8. Florida Department of Environmental Protection , "Information Portal," [Online]. Available: <https://prodenv.dep.state.fl.us/DepNexus/public/searchPortal>. [Accessed 06 05 2026].
9. AECOM Technical Services, Inc, "Comprehensive Utilities Master Plans 2020-2040," 12 December 2023. [Online]. Available: <https://fortmyers.gov/2145/Utilities-Master-Plans>. [Accessed 01 05 2026].
10. Black and Veatch Holding Company, "Capacity Analysis Report for City of Fort Myers Water Treatment Plant," 27 November 2024. [Online]. Available: <https://prodenv.dep.state.fl.us/DepNexus/public/searchPortal>. [Accessed 01 05 2026].
11. Black and Veatch Holding Company, "Capacity Analysis Report for City of Fort Myers Water Treatment Plant," 27 November 2024. [Online]. Available: <https://prodenv.dep.state.fl.us/DepNexus/public/searchPortal>.
12. S. McCoy, *Murach's Python for Data Science (2nd ed.)*, Fresno, CA: Mike Murach and Associates, Inc., 2024.
13. J. F. Adamowski, "Peak Daily Water Demand Forecast Modeling Using Artificial Neural Networks," *Journal of Water Resources Planning and Management*, pp. 119-128, 2008.
14. Liquid Solutions Group, LLC, "Capacity Analysis Report for Toho Water Authority Poinciana Water Supply System," December 2024. [Online]. Available: <https://prodenv.dep.state.fl.us/DepNexus/public/searchPortal>. [Accessed 01 05 2026].
15. Microsoft Corporation, "Microsoft Excel 365," July 2022. [Online].
16. Frontline Systems, Inc, "Frontline Excel Analytic Solvers V25.2.2.0 - Analytical Solvers Data Science (ASDS)," 2025. [Online]. Available: <https://www.solver.com>.
17. Y. Mercadier, "Mean Absolute Error (MAE)," 2024. [Online]. Available: https://distancia.readthedocs.io/en/latest/MeanAbsoluteError.html?utm_source=chatgpt.com. [Accessed 06 05 2026].

18. Faster Capital, "Z score: The Z Score Revelation: Unmasking Outliers in Data," 07 April 2025. [Online]. Available: https://fastercapital.com/content/Z-Score--The-Z-Score-Revelation--Unmasking-Outliers-in-Data.html?utm_source=chatgpt.com. [Accessed 06 05 2026].
19. J. Shu, X. Xia, S. Han, Z. He, P. Ke and B. Liu, "Long-term water demand forecasting using artificial intelligence models in the Tuojiang River basin, China," *PLoS ONE*, vol. 19, no. 5, pp. 1 - 23 <https://doi.org/10.1371/journal.pone.0302558>, 2024.
20. J. Wilber and B. Werness, "The Importance of Data Splitting," 2026. [Online]. Available: <https://mlu-explain.github.io/train-test-validation/>.
21. S. McCoy, *Murach's Python for Data Science (2nd Edition)*, Fresno, CA: Mike Murach & Associates, 2024.
22. Y. P. Jiang, *PowerPoint Slides - Introduction to Predictive Analytics*, 2026.
23. AECOM Technical Services, Inc, "Comprehensive Utilities Master Plans 2020-2040," 12 December 2023. [Online]. Available: <https://fortmyers.gov/2145/Utilities-Master-Plans>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.