4.1.Introduction

In almost studies performed in industrialized countries, the residential water demand function is specified as a single equation linking tap water use (the dependent variable) to water price and a vector of demand shifters (like household socio-economic characteristics, housing features, climatologic variables, etc.) to control for heterogeneity of preferences and other variables affecting water demand (Agthe and Biillings, 1987).

Most of the models that are employed in residential water demand study both in the developed and developing countries are regression models. They typically use the form:

Q=f (P, Z) where P is the price variable and Z are the factors or a range of shifters of demand such as income, household demographics and other characteristics such as weather variables, etc. (Arbuès et al., 2004).

The current chapter presents the proposed methodology for assessing the relationships between WCP and the other parameters by using statistical techniques. The subject of statistics is based around the idea that when it exists a big set of data and the purpose is to analyze that set in terms of relationships between individual points.

The techniques used in this thesis have the following objectives:

- Determine the possible association between the variables.
- Establishing the cause-effect relation between the data set.
- Measuring the strength of association between the variables and the direction of the relationship.
- Create and chose the best model for predicting residential water in the study area.

4.2.General Methodology

The chapter will attempt to give some elementary background mathematical skills that will be required to understand the process of many statistical analyses. These analyses will explore the effect of some parameters on water use. Firstly, the methodology covered mathematical techniques used for the investigation, followed by an explanation about the reason why such technique may be used and what the result of the operation tells about the data. Besides, in the second part an overview of the software, tools, used during the analysis. For more details, multivariate analysis will be conducted to see if there is any predictive relationship between water use and the other factors. The examination of results will be presented in the next chapter of this thesis.

There is a huge variety of machine learning algorithms existing nowadays.

The methodology is composed of many algorithms used for this research and presented in the following steps:

Firstly, to figure out which variables are connected together and to study the strength of a relationship between these continuous variables, **Correlation Analysis** will be used for these purposes followed by **Correlation Matrix**. This last will used to compute the correlation coefficients between variables.

Second step is **ANOVAs Test.** An overview about the use of ANOVA in statistics will be presented. The purpose of this test is to assess any differences in WCP in houses according to every independent variable.

In the third step, **Factor Analysis (FA)** is used, followed by **Principal Components Analysis (PCA)** in the fourth step. This analysis uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Then the fifth step is **Cluster Analysis (CA)**, which divides data into homogeneous and distinct groups (clusters). FA, PCA & CA were uses because they summarize data so that relationships and patterns can be easily interpreted and understood.

The two final steps represent **the Artificial Neural Networks (ANNs) and the Adaptive Neuro Fuzzy Inference System (ANFIS)**. The two techniques were used to configure or to reject the previous tests results, *i.e.,* to assess the main determinants of household water use.

Figure 4.1 presents the proposed methodology used in the present thesis. It attempts to better understand the relationship between household water use and indoor or outdoor factors.

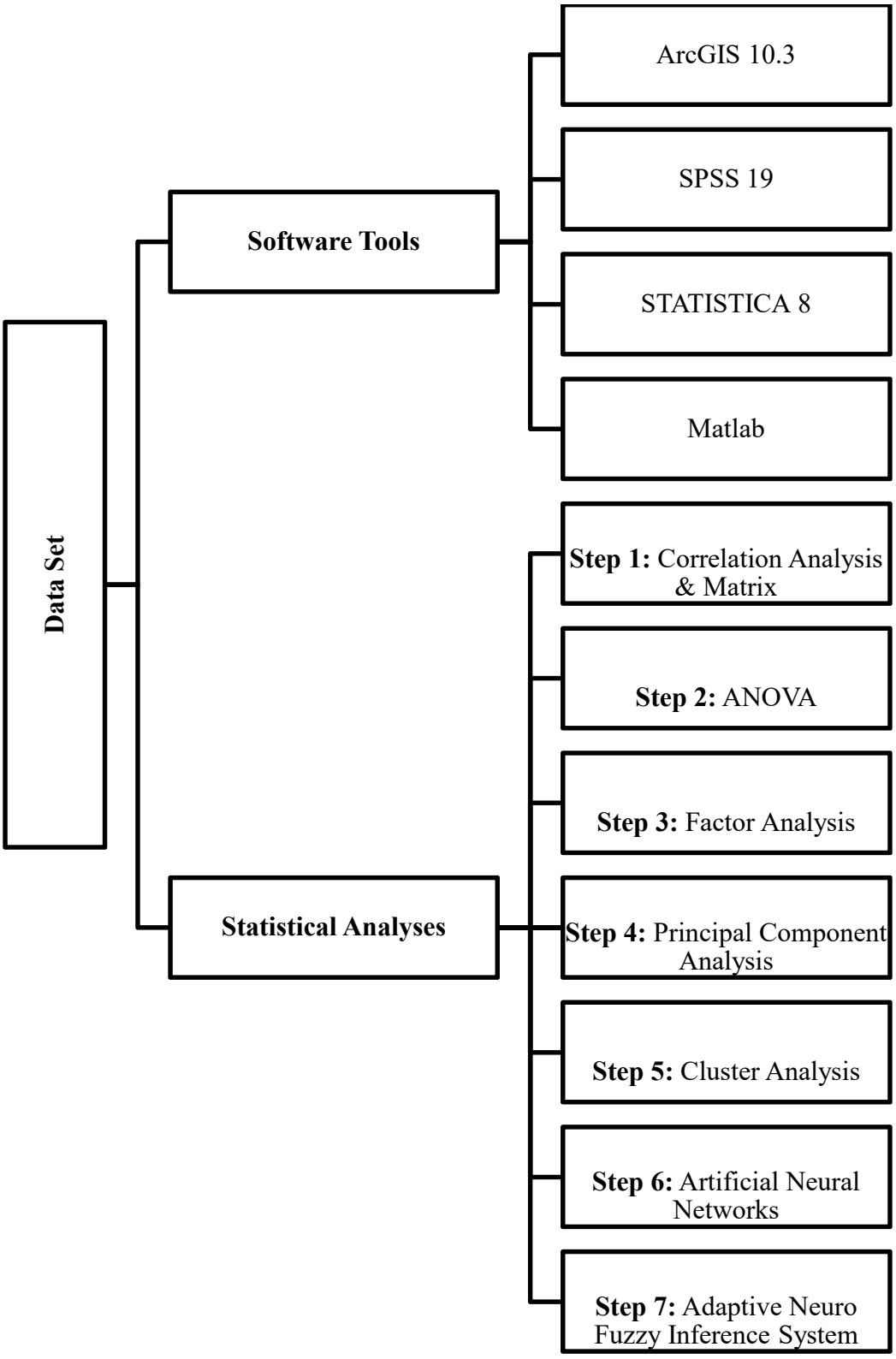

Fig 4.1 Proposed methodology

4.3.Statistical Analysis

4.3.1. Correlation Analysis and Correlation Matrix

The correlation analysis is the statistical technique used to study the closeness of the relationship between two or more variables. The variables are correlated when the movement of another variable accompanies the movement of one variable. The purpose of such analysis is to find out if any change in the independent variable results in the change in the dependent variable or not. Hence, with the correlation analysis the degree of relationship between these variables can be measured in one figure. Usually in statistics, four types of correlations could be measured for this purpose: Pearson correlation, Kendall rank correlation, Spearman correlation and the point-Biserial correlation. In this research the focus will be only on Pearson correlation because is the most widely used.

a) Karl Pearson's coefficient of correlation

The Pearson correlation is widely used mathematical method wherein the numerical expression is used to calculate the degree and direction of the relationship between linear related variables. The coefficient is popularly known as a **Pearsonian Coefficient of Correlation** denoted by **"r"**. If the relationship between two variables X and Y is to be ascertained, then the following formula is used:

$$\boldsymbol{r} = \frac{\sum(\boldsymbol{X}-\bar{\boldsymbol{X}})(\boldsymbol{Y}-\bar{\boldsymbol{Y}})}{\sqrt{\sum(\boldsymbol{X}-\bar{\boldsymbol{X}})^2}\sqrt{(\boldsymbol{Y}-\bar{\boldsymbol{Y}})^2}} \qquad \textbf{(4.1)}$$

Where: $\bar{X}$ is the mean of X variable and $\bar{Y}$ is the mean of Y variable. The value of "r" is always lies between$\pm\mathbf{1}$, such as:

r = +1 indicates perfect and positive correlation.

r= -1 indicates perfect and negative correlation.

r= 0 means no correlation.

Regression analysis is widely used for prediction and forecasting, where use has substantial overlap with the field of machine learning. It is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. The degree of which the variables are correlated to each other depends on the Regression Line, *i.e.* all the points plotted are connected via a line in the manner that the distance from the line to the points is the smallest. The differences between correlation and regression are:

- The correlation coefficient measures the "degree of relationship" between variables, X and Y whereas the regression analysis studies the "nature of relationship" between variables.

- Correlation coefficient does not clearly indicate the cause/effect relationship between the variables, i.e. it cannot be said with certainty that one variable is the cause, and the other is the effect. Otherwise, the regression analysis clearly indicates the cause-effect relationship between the variables.

b) Regression line

The regression line is the line that the best fits the data, such that the overall distance from the line to the points (variable values) plotted on a graph is the smallest. In other words, a line used to minimize the squared deviations of predictions is called as the regression line.

c) Regression equation

The regression equation is the algebraic expression of the regression lines. It is used to predict the values of the dependent variable from the given values of independent variables. It can be expressed as follows:

$$\mathbf{Y = a + b\,X} \qquad \mathbf{(4.2)}$$

Where Y is the dependent variable (or sometimes, the outcome, target or criterion variable), X is the independent variable (or sometimes, the predictor, explanatory or regressor variables), a and b are the two unknown constants that determine the position of the line.

d) Assumptions

- **Assumption 1:** the dependent variable should be measured on a continuous scale (i.e. measured in hours, in Kg…etc).
- **Assumption 2:** the existence of two or more independent variables.
- **Assumption 3:** the existence of independence of variations (i.e. independence of residuals), which can be easily checked by using the Durbin-Watson statistics.
- **Assumption 4**: linear relationship between (a) the dependent variable and each of your independent variables, and (b) the dependent variable and the independent variables collectively. Using scatterplot and partial regression plots can check these relationships.
- **Assumption 5:** the data needs to show homoscedasticity, which is where the variance along the line of best fit remains similar as the moving along the line.
- **Assumption 6:** the data must not be multicollinearity, which means the existence of two or more correlated independent variables with each other.
- **Assumption 7:** no significant outliers in the data.
- **Assumption 8:** the dataset must be approximately normally distributed (i.e. fit the shape of a bell curve).

4.3.2. ANOVA Test

The acronym ANOVA refers to **analysis of variance** and is a statistical procedure used to test the degree to which two or more groups vary or differ in an experiment. In other words, they help to figure out if there is a need to reject the null hypothesis or accept the alternate hypothesis. Basically, a null hypothesis is the assumption that there will be no differences between groups that are tested. There are two main types of ANOVA, one-way and two-way test. One-way and two-way refers to the number of independent variables in the dataset.

a. One-way ANOVA

Used between two groups to see if there's a difference between them. It used to compare two means from two independent groups using the f-distribution. The null hypothesis for the test is that the two means are equal. Therefore, a significant result means that the two means are unequal.

b. Two-way ANOVA

Used when you have two groups, *i.e.* has two independent variables (can have multiple levels). Levels are different groups in the same independent variable.

The f value in ANOVA is a tool to help answering the question "is the variance between the means of two populations significantly different?" The f ratio is a test statistic, calculated as:

F value = variance of the group means (Means square between)/ mean of the within group variances (Mean squared error).

c. Assumptions

- The population must be closer to normal distribution
- Samples must be independent
- Population variances must be equal
- Groups must have equal sample sizes

4.3.3. Factor Analysis (FA)

In this part, a method will be used to restructure the dataset by reducing the number of variables; often called "data reduction" or "dimension reduction" technique. The way that the information contained is measured is by considering the variability within and co-variation across variables that is the variance and co-variance (correlation).

Either the reduction might be by discovering that a particular linear compensation of the variables accounts for a large percentage of the total variability in the data or by discovering that several of the variables reflect another "latent variable".

Broadly the process used three ways:

- To discover the linear combinations that reflects the most variation in the data.
- To discover if the original variables are organized in a particular way reflecting another "latent variable" (called exploratory factor analysis_ EFA).
- To confirm a belief about how the original variables are organized in a particular way.

The factor analysis is a technique of clumping subgroups of variables together based on their correlations and often just by looking at the correlation matrix and spotting clusters of high correlations between groups of variables can show what the factors are going to be. (Note that: Latent variable= construct= factor).

a. Perform factor analysis

In some type of statistic associated probability density function to produce a *p value*. Two such statistics are **the Bartlett test of Sphericity** and **the Kaiser-Meyer-Olkin** measure of sampling adequacy (usually called the **MSA**).

The Bartlett test of Sphericity compares the correlation matrix with a matrix of zero correlations (called the identity matrix). It is used to reject or not the hypothesis according to which the variables are not correlated. While, The MSA does not produce a p value but some researchers like Norman & Streiner in p198 recommend that it's better to remove variables with an MSA below 0.7.

b. Extracting the factor analysis

Principal components and Principal axis are two factoring extraction methods. PCA is not a type of factor analysis but it gives very similar results.

c. Rotation for factor analysis

Rotations can be applied on the factors. Several methods are available including Varimax, Quartimax, Promax, etc.

4.3.4. Principal Component Analysis (PCA)

PCA is a powerful and popular multivariate analysis method that can identify patterns in datasets with quantitative variables, and expressing the data in such a way as to highlight their similarities and differences. The aim of this technique is the extraction of the important information from this data set and representation of it with a set of new orthogonal variables that are called as principal components (PCs) (Abdi, et al., 2010).

a) The process of conducting a PCA

- The inter-correlations amongst the items are calculated yielding a correlation matrix.
- The inter-correlated items or "factors" are extracted from the correlation matrix to yield "principal components".
- These "factors" are rotated for purposes of analysis and interpretation.

b) Eigenvalues and inertia

Eigenvalues are the amount of information (inertia) summarized in every dimension. The first dimension contains the highest amount of inertia. The number of Eigenvalues is equal to the number of non-null Eigenvalues.

c) The correlation circle or variables chart

The correlation circle shows the correlations between the components and the initial variables. The variables can be displayed in the shape of vectors.

d) The Biplots

The Biplots represent the observations and variables simultaneously in the new space. There are different types of Biplots; correlation biplot, distance biplot and symmetric biplot.

e) Assumptions for conducting FA, PCA and CA

- A large enough sample size to allow the correlations to converge into mutually exclusive "factors".
- Normality and linearity of dataset
- The sample must be relatively homogeneous
- No significant outliers

4.3.5. Cluster Analysis (CA)

One of the useful analytical tools is unsupervised clusterization, where the data is classified by the algorithm into specified number of classes based on internal patterns. It can be used to search for the subtypes and subclasses for researched process, value or compound (Likas et al., 2003). There are a number of different methods that can be used:

a. Hierarchical methods

- *Agglomerative methods*: in which subjects start in their own separate cluster. The two 'closest' (most similar) clusters are then combined and this is done repeatedly until all subjects are in one cluster. At the end the optimum number of clusters is then chosen out of all cluster solutions.
- *Divisive methods:* in which all subjects start in the same cluster and the above strategy is applied in reverse until every subject is in a separate cluster.

b. Non-hierarchical methods (Known as K-means clustering methods)

In this work the concentrate will be on the former rather than the latter.

When carrying out a hierarchical cluster analysis, the process can be represented on a diagram known as a Dendrogram. This diagram illustrates which clusters have been joined at each stage of the analysis and the distance between clusters at the time of joining. If there is a large jump in the distance between clusters from one stage to another then this suggests that at one stage clusters that are relatively close together were jointed whereas, at the following stage, the cluster that were jointed were relatively far apart. This implies that the optimum number of clusters may be the number present just before that large jump in distance. The data is classified firstly by setting k centroids, which will be the core to the searched classes. Then, the grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. At each of iteration cluster center is recalculated until the best position is reached (Likas et al., 2003).

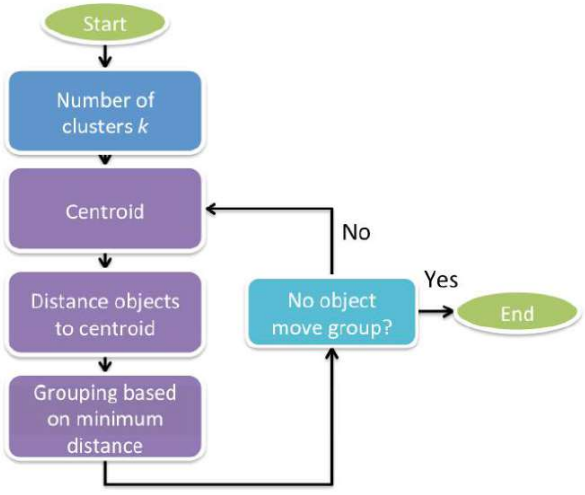

Fig 4.2 K-means clustering algorithm (Likas et al., 2003)

4.3.6. Artificial Neural Networks (ANNs)

This part of the chapter introduces brief information about artificial neural networks. Over the past, there has been an explosion of interest in neural networks. It started with successful application of this powerful technique across a wide range of problem domains, in areas as diverse as finance, medicine, engineering, geology and even physics. The artificial neural networks (ANNs) are one of the most popular machine learning models nowadays. It was first introduced in 1943 (McCulloch and Pitts, 1943) research into application of ANNs has blossomed since the introduction of back propagation training algorithm for feed forward ANNs in 1986 (Rumelhart *et al.*, 1985).

The basic of this model is in several layers that are made up of a number of interconnected nodes, containing the activation function. Data is fed into the input layer (independent) and is transformed by weights and neurons through the hidden layers (one or more hidden layers), then sent to the output layer (dependent). Produced output is sent to outer world as result. ANNs have different types such as Multilayer Perceptron Neural Networks (MLP) and General Regression Neural Network (GRNN). In this paper, MPL model was adopted since its training pattern propagates several times until a lower error is achieved.

a. General Characteristic of ANNs

More recently, artificial intelligence techniques (AI) are adopted increasingly in view of their ability to do non-linear curve fitting and applicability to a very complex data set, it used where the structure of the model is unknown and dealing with noisy data (Tiwari and Adamowski, 2013). Additionally, "Soft Computing" such as Artificial Neural Networks (ANNs) and Fuzzy Logic (FL) are more efficient and less time consuming in modeling complex systems (Pahlavan *et al.*, 2012; Sonmez *et al.*, 2018).

ANNs are adaptive where they take data and learn from it. They can reduce development time by learning underlying relationships even if they are difficult to find and describe. They can also solve problems that lack existing solutions. ANNs can generalize, where they can correctly process data that only broadly resembles the data they were trained on originally. Similarly, they can handle imperfect or incomplete data, providing a measure of fault tolerance. Generalization is particularly useful in practical applications because real world data is noisy. ANNs can capture complex interactions among the input variables in a system. They are highly parallel, i.e., their numerous identical, independent operations can be executed simultaneously (Aggrawal and Song, 1997).

b. Disadvantages of ANNs

Like any approach, ANNs has also disadvantages. They can be difficult to account for their results. ANNs are like human experts and express opinions that they cannot easily explain. In addition, training methods are imperfectly understood, where few definite rules exist for choosing the optimum architecture and there is no definite way of finding the best solutions which also depend in practice on the accuracy of the training data used. They can consume large amounts of computer time, especially during training (Aggrawal and Song, 1997).

c. Basic Components of ANNs

The basic of ANNs model is in several layers that are made up of a number of interconnected nodes, containing the activation function. The training set is presented to a model through input layer; one or more hidden layers perform processing by the system of weighted connections, taking each of the inputs for calculation and finally output gives the fitted function (Cheng and Titterington 1994).

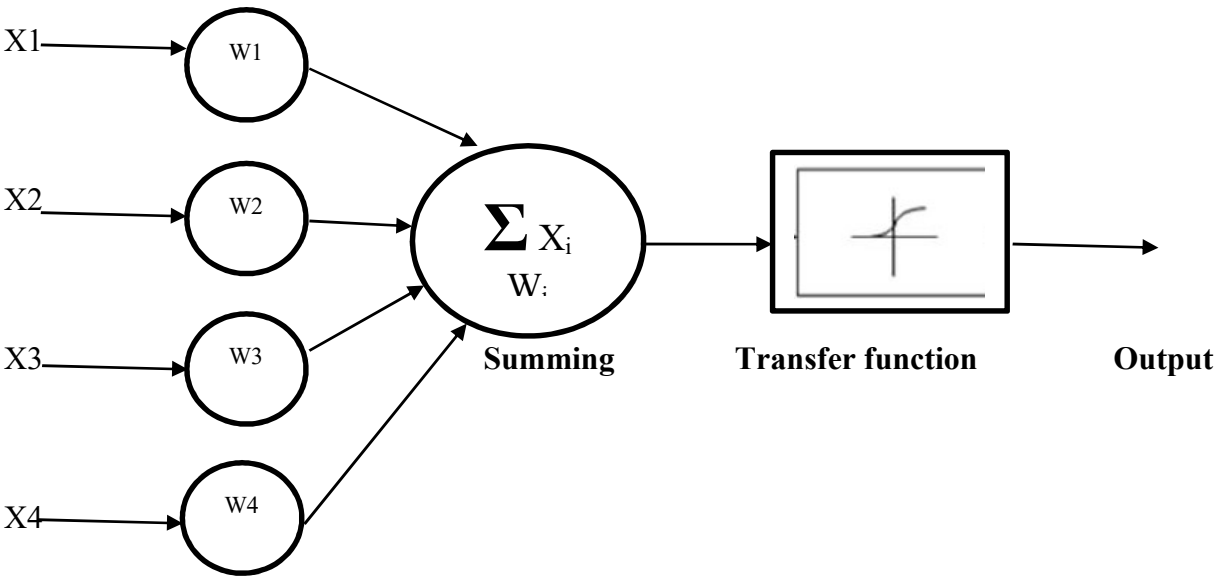

Fig 4.3 Schematic representation of an artificial neuron

The components of ANNs are:

- **Inputs**

The inputs (X_1, X_2, X_3… X_i) are elements of ANNs, which take data from outer world. They do not have any other functions than transform data to the next step. A neuron can have unlimited number of inputs but, there must be only one output of every neuron.

- **Weights**

The weights (W_1, W_2, $W_3 \ldots W_i$) are most important elements of mathematical neuron by itself and ANNs generally. Because learned data by the networks is stored on the weights. Accordingly, network's synaptic weight vector expressing as $W= [W_i]_{nx}$ consists convertible values. Typically, initial values of weights are selected as a random value in (-1, 1) range. How to set up of a weight for learning relationship between the given variables is decided based on selected learning rule. Weights of ANNs can be thought as synapse in biological nervous system.

- **Summation Function**

The summation function is responsible for summation of all data coming from outer world and related weights. It is shown in equation (5.3):

$$\mathbf{vi} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{xi}.\,\mathbf{wi} - \boldsymbol{\theta} \qquad \textbf{(4.3)}$$

Summation function transfers the created weighted input (v_i) to activation function. (θ) Value expresses the threshold value. Use of threshold value in summation function is not obligatory. Some other functions that can be used of summation function are given in the following table.

Table 4.1 Some functions can be used instead of summation function

Functions	Weights input (v_i)
Multiplication	$v_i= \prod_{i=1}^{n} xiwi$
Maximum	$v_i= \max (x_i\ w_i)$
Minimum	$v_i= \min (x_i\ w_i)$
Signum	$v_i= \text{sgn} (x_i\ W_i)$

- **Activation (Transfer) Function**

The activation function is responsible for activation of coming weighted input (v_i) and determination of the final output value. It varies based on type of problem and there is no a universal formula for which type of activation function should be used. The choice of activation function depends largely on available data and what the designed learning of network. Sigmoid and hyperbolic tangent functions are most commonly used. The equation below is the mathematical expression of activation function:

$$\mathbf{f}(\mathbf{v_i}) = \mathbf{y} \qquad \textbf{(4.4)}$$

- **Output Function**

The output function is responsible for transfer of output value of activation function to outer world as network's final output value or to other connected neuron as their input values.

The artificial neuron components come together in three layers and parallel to each other for forming ANNs, not randomly, as follow:

- **Input layer:** Neurons transfer the information coming from outer world to hidden layers.
- **Hidden layers:** Information coming from input layer is processed and sent to output layer. There can be more than one hidden layer.
- **Output layer:** In this final layer, process elements. Transferred information from hidden layers are handled and produce output for given data in input layer. Produced output is sent to outer world as result (Lippmann, 1987). Layers are connected together with weights, as shown in the next figure.

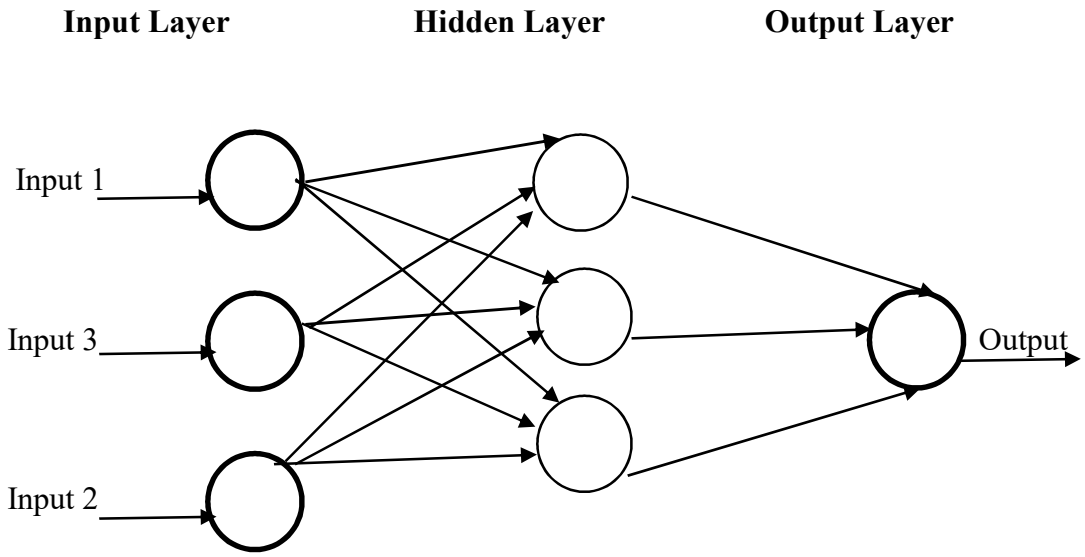

Fig 4.4 Artificial neural network representation with 3 layers

d. Types of ANN Based on Structure

ANNs are divided into two types: Feed Forward ANNs and Back Forward ANNs.

- **Feed Forward Networks**

In this group, neurons are located into layers generally. Inputs are sent from one layer to next one by one-way weights. Due to the one-way connection, a return back of the next layer's output to previous layer as input is impossible. Multi-layer Perceptron (MLP) and Learning Vector Quantization (LVQ) networks are examples for feed forward networks.

- **Feed-Back Networks**

In this group, outputs of hidden and output layer are fed to input layer or previous hidden layers. In this way, inputs can be transferred in both forward and reverse directions. This type networks have dynamic memory and an output in a moment reflects inputs at that moment and previous inputs.

They are suitable for prediction applications and quite successful in the estimation of various types of time series. Examples of these networks are SOM (Self Organization Map), Elman and Jordan networks.

e. Selection of Learning and ANNs Structure

The selection of appropriate ANNs structure depends on considered learning algorithm. In the next table, the differences between network categories and the common network structures.

Table 4.2 Network type and their intended use (Anderson, et al., 1992)

Indented Use of Network	Networks	Use of Network
Prediction	▪ Back-propagation ▪ Delta bar delta ▪ Extended delta bar delta ▪ Directed random search ▪ Higher order neural networks ▪ Self-organizing map (SOM) into back-propagation	Using of inputs values for prediction of some output
Classification	▪ Learning vector quantization (LVQ) ▪ Counter-propagation ▪ Probabilistic neural networks	Using of inputs for classification
Data Association	▪ Hopfield ▪ Bolzmann machine ▪ Hamming network ▪ Bidirectional associative memory ▪ Spatio-temporal pattern recognition	Determining of incorrect values and completing of missing values in input data
Data Conceptualization	▪ Adaptive resonance network (ART) ▪ Self-organizing map (SOM)	Analyze of for derivation of grouping relationships
Data Filtering	▪ Recirculation	Smooth of an input signal

f. Identification of Performance Function

Performance functions calculate the cumulative values between the target outputs values and created outputs by the network. According to these calculated values, how the network close to the pattern of training set is observed and using these values changes connection weights. Mean square error (MSE) is commonly used performance function in feed forward networks.

$$\mathbf{MSE}=\frac{1}{n}\sum_{i=1}^{n}[\mathbf{e}\ (\mathbf{t})\ ^{2}] \qquad \mathbf{(4.5)}$$

Mean Error (ME), Root Mean Square (RMSE) and Mean Absolute Error (MAE) are some of the other performance functions can be used, their equations are:

$$\mathbf{ME}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}\ (\mathbf{t}) \qquad \mathbf{(4.6)}$$

$$\mathbf{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}[\mathbf{e}(\mathbf{t})^{2}]} \qquad \mathbf{(4.7)}$$

$$\mathbf{MAE}=\frac{1}{n}\sum_{i=1}^{n}|\mathbf{e}(\mathbf{t})| \qquad \mathbf{(4.8)}$$

$$\mathbf{e_t}=\frac{y_t-t_t}{t_t}\times 100 \qquad \mathbf{Then\ MAE}=\frac{1}{n}\sum_{i=1}^{n}|e_t| \qquad \mathbf{(4.9)}$$

Where; e (t) is forecast error at period t, n is number of periods, y_t is forecast and t_t is actual result at period t (Agami *et al*., 2009).

4.3.7. Fuzzy Logic (FL)

Fuzzy logic is primarily concerned with the manipulation of sets with non-deterministic boundaries (Zadeh, 1994). In this regard, fuzzy logic can be used for approximate reasoning. Since in forecasting one usually deals with uncertain outcomes, fuzzy models are one class of widely used tools in this area. Fuzzy logic is especially useful in modeling qualitative and imprecise data and systems, such as human reasoning processes, making decisions based upon vague or imprecise data and considering uncertainty at various levels (Zadeh, 1994 and Iyatomi & Hagiwara, 2004).

4.3.8. Adaptive Neuro Fuzzy Inference System (ANFIS)

Along with fuzzy logic and models, the hybridization of fuzzy logic and ANNs, resulting in neuro-fuzzy models, has also been widely used in water demand forecasting. Neuro-fuzzy models possess the pattern recognition ability of ANNs and the reasoning ability of fuzzy logic (Nauck & Kruse 1997).

Fuzzy logic is employed to describe human thinking and reasoning in a mathematical framework. The main problem with fuzzy logic is that there is no systematic procedure to define the membership function parameters. The construction of the fuzzy rule necessitates the definition of premises and consequences as fuzzy sets. On the other hand, an ANNs has the ability to learn from input and output pairs and adapt to it in an interactive manner (Yurdusev et al., 2009). In recent years, the ANFIS method, which integrates ANNs and FL methods, has been developed. ANFIS has the potential benefits of both these methods in a single framework. ANFIS eliminates the basic problem in fuzzy system design, defining the membership function parameters and design of fuzzy if-then rules, by effectively using the learning capability of ANNs for automatic fuzzy rule generation and parameter optimization (Nayak et al., 2004).

In the present work, the ANFIS methodology is proposed to self-organize model structure and to adapt parameters of the fuzzy system water consumption prediction and rules governing water consumption in the region. It has the advantage of allowing the extraction of fuzzy rules from numerical data. The main drawback of the ANFIS prediction model is the time requested for training structure and determining parameters (Chang & Chang, 2006 and Sen, 2001).

a. ANFIS Components

Fuzzy inference system (FIS) is a rule-based system consisting of three components. These are: A rule-base, containing fuzzy if-then rules, a data-base, defining the membership functions (MF) and an inference system that combines the fuzzy rules and produces the system results (Sen, 2001).

b. ANFIS Structure

- The first phase of FL modeling is the determination of membership function (MF) for input and output variables.
- The second phase is the construction of fuzzy rules.
- The last phase is the determination of output characteristics, output MF and model results (Yurdusev et al., 2009). A general structure for FIS is shown in figure below.

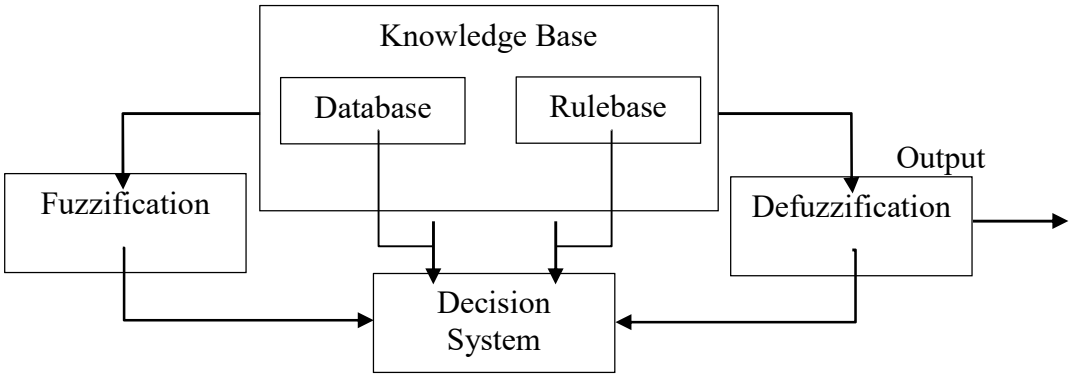

Fig 4.5 The general structure of the fuzzy inference system (Yurdusev et al., 2009)

- **Fuzzification** is a way of mapping numeric input variables into linguistic fuzzy sets. Fuzzification is a mathematical procedure for converting an element in the universe of discourse into the membership value of the fuzzy set. Universe discourse includes a system consisting of a number of input and output variables (Yurdusev et al., 2009).
- **Defuzzification** converts the conclusions of the inference mechanism into numerical values. The aim of defuzzification is always to transform a fuzzy set into a crisp number (Yurdusev et al., 2009).

c. ANFIS types

There are two types of FISs, described in the literature, Sugeno-Takagi FIS and Mamdani FIS. In this study, Sugeno Takagi FIS model is used for predicting water consumption.

The most important difference between these systems is definition of the consequence parameter. The consequence parameter in Sugeno FIS is either a linear equation, called ‘‘first-order Sugeno FIS”, or constant coefficient, ‘‘zero-or- der Sugeno FIS (Jang et al., 1997).

d. ANFIS Process

ANFIS uses the learning ability of ANNs to define the input–output relationship and the fuzzy rules are constructed by determining the input structure. The system results are obtained through the reasoning capability of FL (Yurdusev et al., 2009).

The steps of conducting the process are:

- Generate the initial FIS model by choosing one of the following partitioning techniques: Grid partition (Generates a single-output Sugeno-type FIS by using grid partitioning on the data) or Sub. clustering (Generates an initial model for ANFIS training by first applying subtractive clustering on the data).
- There are two methods that ANFIS learning employs for updating membership function parameters: Backpropagation for all parameters (a steepest descent method) A hybrid method consisting of backpropagation for the parameters associated with the input membership functions, and least squares estimation for the parameters associated with the output membership functions.
- Training Error: The training error is the difference between the training data output value, and the output of the fuzzy inference system corresponding to the same training data input value, (the one associated with that training data output value). The training error records the root mean squared error (RMSE) of the training data set at each epoch.
- Checking Error: The checking error is the difference between the checking data output value, and the output of the fuzzy inference system corresponding to the same checking data input value, which is the one associated with that checking data output value. The checking error records the RMSE for the checking data at each epoch.

4.4.Software and Tools

The second part of this chapter gives more details about the tools used in this research. The used statistical software are : ArcGIS, SPSS, STATISTICA and MATLAB Software.

4.4.1. Arc GIS 10.3

ArcGIS is a geographic information system (GIS) for working with maps and geographic information. It is designed for creating and using maps, compiling geographic data, analyzing mapped information, sharing and discovering geographic information, using maps and geographic information in a range of applications and mapping geographic information in a database. The system provides an infrastructure for making maps and geographic information available throughout an organization, across a community. Esri released ArcGIS version 10.3 in December 2014 that used in this research.

In the present paper, ArcGIS is used to realize the thematic analysis, to illustrate the variation of WCP through the years, to demonstrate the variation of the socio-economic variables, the physical characteristics of housing units and the indoor habits of residents in the study area.

4.4.2. Statistical Package for the Social Sciences (SPSS 19)

The software name originally stood for Statistical Package for the Social Sciences (SPSS) reflecting the original market, although the software is now popular in other fields as well. SPSS statistics is software package widely used for statistical analysis. Market researchers, health researchers, survey companies, government, education researchers, and others use it.

The SPSS Statistic version 19 is used in this research. Statistical input has a two-dimensional table structure, where the rows represent cases (such as individuals or households) and the columns represent measurements (such as age, gender, household income). Only two data types are defined: *numeric and text*.

The first procedure or step in SPSS is the creation of the dependent and the independent variables. Output results are varying, according to how many variables exist, from tables, graphs like scatterplots, histograms, pie charts, etc. The SPSS software can conduct the statistical description of variables, regression analysis, ANOVA test, etc.

4.4.3. STATISTICA 8

STATISTICA is statistical software used around the world in more than 20 countries. Stat Soft STATISTICA line of software has gained unprecedented recognition by users and reviews. STATISTICA is a software package that offers much to both students of and professionals. It has a friendly graphical user interface, make it easy to be learnt, in addition to its extensive help facility.

It produces statistical results in high quality output format, seamlessly integrating with several Word-processor and Spreadsheet programs, under the Microsoft Windows environment. It has limitations, through, but these are more likely to be felt by advanced users and statisticians who need fine control over the computational process (STATISTICA, 2011).

With the available water consumption, socio-economic, physical and meteorological data during the period of study, the relationship between water consumption and the independent variables can be defined by many statistical technique

s. STATISTICA 8 can provide lots of statistical analysis including: basic, multiple regressions, ANOVA, principal component analysis, cluster analysis, nonparametric and distribution fittings among other statistics. Also, it enables statisticians and researchers to conduct data mining, in order to check their model's accuracy. One of the perfect abilities of this statistical software is the high-quality graphs, where it is able to produce histograms, scatter plots, mean plot, box plot, and availability and line graphs.

4.4.4. MATLAB

All network training and testing processes in this thesis was performed in MATLAB, which provides a solid ANNs toolbox. This toolbox has built-in functions to construct, train, and save the network.

Matlab is a multi-paradigm numerical computing environment and proprietary programming language developed by MathWorks. Matlab allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages.

The application of the whole analysis on dataset, besides to the presentation of results and their interpretation is presented with details in following chapters.