*Article*

# Short-Term and Long-Term Forecasting for the 3D Point Position Changing by Using Artificial Neural Networks

Eleni-Georgia Alevizakou [1], George Siolas [2] and George Pantazis [1,*]

[1] School of Rural and Surveying Engineering, National Technical University of Athens, 15780 Zografos, Athens, Greece; elalevizakou@gmail.com
[2] School of Electrical and Computer Engineering, National Technical University of Athens, 15780 Zografos, Athens, Greece; gsiolas@islab.ntua.gr
* Correspondence: gpanta@central.ntua.gr; Tel.: +30-210-772-2696

**Abstract:** Forecasting is one of the most growing areas in most sciences attracting the attention of many researchers for more extensive study. Therefore, the goal of this study is to develop an integrated forecasting methodology based on an Artificial Neural Network (ANN), which is a modern and attractive intelligent technique. The final result is to provide short-term and long-term forecasts for point position changing, i.e., the displacement or deformation of the surface they belong to. The motivation was the combination of two thoughts, the insertion of the forecasting concept in Geodesy as in the most scientific disciplines (e.g., Economics, Medicine) and the desire to know the future position of any point on a construction or on the earth's crustal. This methodology was designed to be accurate, stable and general for different kind of geodetic data. The basic procedure consists of the definition of the forecasting problem, the preliminary data analysis (data pre-processing), the definition of the most suitable ANN, its evaluation using the proper criteria and finally the production of forecasts. The methodology gives particular emphasis on the stages of the pre-processing and the evaluation. Additionally, the importance of the prediction intervals (PI) is emphasized. A case study, which includes geodetic data from the year 2003 to the year 2016—namely X, Y, Z coordinates—is implemented. The data were acquired by 1000 permanent Global Navigation Satellite System (GNSS) stations. During this case study, 2016 ANNs—with different hyper-parameters—are trained and tested for short-term forecasting and 2016 for long-term forecasting, for each of the GNSS stations. In addition, other conventional statistical forecasting methods are used for the same purpose using the same data set. Finally the most appropriate Non-linear Autoregressive Recurrent network (NAR) or Non-linear Autoregressive with eXogenous inputs (NARX) for the forecasting of 3D point position changing is presented and evaluated. It is proved that the use of ANNs, in order to make short-term and long-term forecasts, provides forecasting changes of the order of 2 mm with Mean Absolute Error (MAE) of the order of 0.5 mm.

**Keywords:** Artificial Neural Network (ANN); non-linear autoregressive recurrent network (NAR); non-linear autoregressive with eXogenous inputs (NARX); Geodesy; 3D position changing; short-term forecasting; long-term forecasting; Data Mining

## 1. Introduction

In the 75 years since the introduction of ANNs basically in Neuroscience [1], their use has expanded to numerous fields, including Economics, Mathematics, Meteorology, Clinical Medicine, Environmental Area etc. Specifically in recent years, ANNs have found a number of applications in the area of recognition and classification [2], water quality [3,4], meteorology [5], politics [6], medical

diagnostics [7] etc. Therefore, it appears that the ANNs are used in applications for classification, clustering, association, regression, modeling and forecasting.

Moreover, ANNs have also been introduced to solve various problems in the scientific field of Geodesy. They have been used in the regional mapping of the Geoid [8], sea level prediction [9], coordinate transformation [10], Coordinated Universal Time (UTC) correction [11,12] etc.

In brief, an ANN is an artificial intelligence technique that mimics the human brain's biological neural network. It is defined as a system of simple processing elements called "artificial neurons", which are typically organized in layers, so that each ANN includes: an input layer, hidden layers (which can have neurons linked in different ways to the other hidden layers or to the output layer) and an output layer. The main feature of ANNs is their inherent capacity to learn, a key component of their intelligence. Learning is achieved through training, a repetitive process of gradual adaptation of the network's parameters to the values required to solve a problem [13–19].

Before analyzing the proposed forecasting methodology it is essential to mention, the worldwide used forecasting techniques. The techniques, which are used in order to produce a forecast, can be separated in two main categories: the Judgmental Forecasting Techniques and the Statistical Forecasting Techniques [20]. However, there is another alternative way of separate these techniques in: the Judgmental ones, the Extrapolative, the Econometric and finally the Non-Linear Computer-intensive ones.

Finally, they can also be distinguished, according to the method which is used in order to produce the forecast, in the following categories: Quantitative forecasting (as Timeseries methods, Causal Relationship or Explanatory methods and Artificial Intelligence methods), Qualitative or Judgmental forecasting (as Individual methods and Committee methods) and Technological forecasting (as Exploratory methods and Normative methods). Also the above mentioned techniques can further split in their corresponding subcategories [21], e.g., the Timeseries methods consist of exponential smoothing method, Kalman Filters [22,23], Box Jenkins ARIMA methods, multiple regression methods, and so on.

After the necessary bibliographic research, it was found that in the scientific field of Geodesy the main interest of the Geodesists is focused on monitoring of a phenomenon rather than on its forecasting. Thus, it was emerged that there is no methodology, which could be used in order to forecast the change of a point position on an artificial structure or on the earth surface. The majority of the techniques or models are focused on the monitoring a point's position.

The aim of this paper is to present an automatic and integrated methodology in order to forecast the position of a point, which belongs to a construction or to the earth surface, in a future time. This forecast concerns both the long-term future as well as the short-term.

The use of the term "forecasting" is more accurate than the term "prediction", because the first one is more objective rather than the second, which is more subjective and intuitive. According to Lewis-Beck "*forecasting can be seen as being a subset of prediction, with all forecasts being predictions, but not all predictions being forecasts*" [24].

The proposed methodology is based on the combination of two disciplines: the Science of Knowledge Discovery in Databases (KDD) for Data Mining and the subset of the Artificial Intelligence Science, namely the ANNs.

Essentially, KDD is a science that emphasizes in the data analysis process, but it does not deal with the methods which will be used to perform this analysis [25–27] and ANN is considered as one of the modern mathematical—computational methods which are used to solve a large number of different kinds of problems.

## 2. Description of the Methodology

The first step is the definition of the examined forecasting problem. This means that the aim of the forecasting and the future use of the results need to be clear. In addition, it is useful to define the

timescale (one-step-ahead forecasting or multi-step-ahead forecasting, namely short-term or long term) of the desired forecasting.

The problem, which the methodology deals with, is the forecasting of the position of a point in a future time. It is important to mention that the desired forecasting depends on the available data set. No change in the position of a point of the order of a few mm can be forecasted if no such displacement is recorded in the existing data, i.e., if the geodetic instruments and methods, which collect the data, cannot detect such a displacement.

Thus the basic steps of the methodology are: Problem definition → finding the available data in time series form → preliminary data analysis (data pre-processing) → ANN definition → evaluation → production of forecasts. These steps are presented in Figure 1.
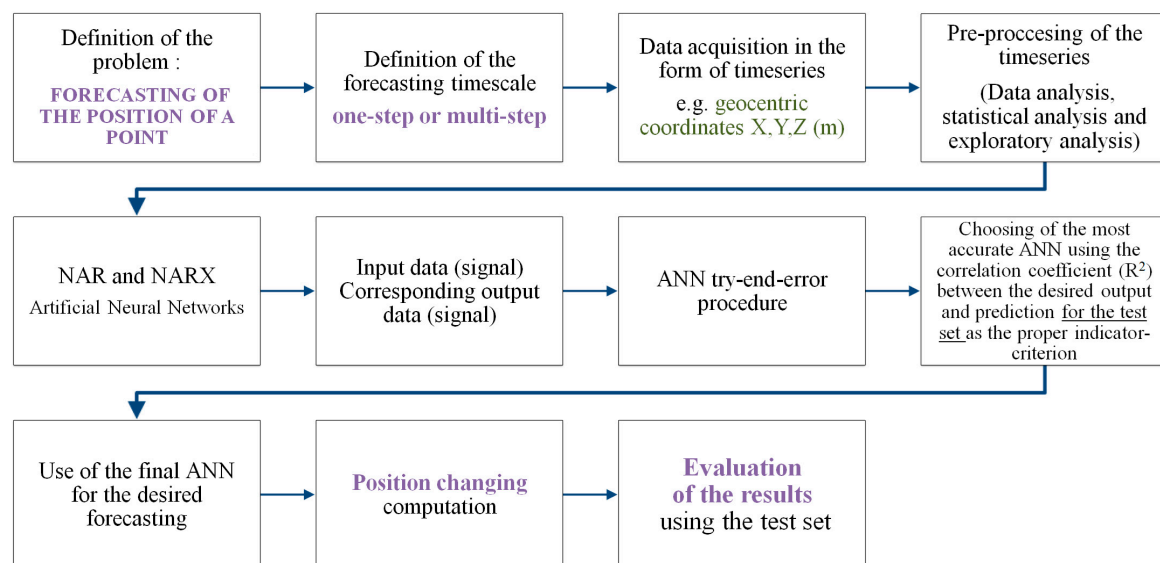


**Figure 1.** Flow chart diagram of the proposed forecasting methodology.

## 2.1. Preliminary Data Analysis (Data Pre-Processing)

This stage is of a great importance since it has a significance role in the accuracy of the results. Through this process the mechanism that produces the timeseries can be found.

The first test is to identify if the timeseries follows a known distribution. This test can be performed either by the Kolmogorov-Smirnov test [28] or its variation, the Anderson-Darling Gof (goodness of fit) test [29]. If the timeseries does not follow a known distribution, then it is possible to transform it. The transformation that can be used for this purpose is the Box-Cox [30,31], so that the abnormal timeseries becomes almost normal. Johnson transformation can also be used as an alternative method [32].

Additionally, the statistical analysis (statistics) of the timeseries is then used to find statistical measures. The computational measures, which are used, are: the mean, the maximum and minimum values, the standard deviation, the covariance, the autocovariance, the correlation and the autocorrelation.

The next step is to find the qualitative features and the timeseries analysis of the major components. At this stage is checked if the timeseries is stationary or have any trend, seasonality, cyclicality or randomness. To convert the timeseries from no stationary to stationary, the diffusion method is used to calculate either the first or the second differences.

Finally, it is sometimes useful to analyze the frequency of the timeseries in order to find some information that is not immediately apparent in the original signal (timeseries). Then there is the possibility to use some mathematical transformations such as the Fourier Transformation, the Fourier

Short-Time Transformation and the Wavelet Transformation, with the ultimate goal of retrieving some additional useful information.

It is then necessary to manage:

- the missing values presented in the timeseries,
- the possible zero values in the event that it is known that they should not exist,
- the duplicate recordings that may exist and
- any unusual observations (outliers) [33].

With regard to the missing values likely to be presented in the timeseries, if they involve long periods of incomplete data, an effort should be made to find these values from other sources. If there are individual cases (not continuous time) and when no seasonal behavior has been observed from the plot of the timeseries, this missing value (or an unexpected zero value) is defined as the mean of the previous and the next value. In the case of duplicate recordings, they must be detected and deleted.

The most complex case is the existence of outliers. The difficulty of this case lies in detecting these values not in an optical way but in a mathematical way that can be automated. Once those individual values are found, they are replaced by the method used in the previous cases, if that is possible. A definition of the outliers is given as follows [34]: "*An outlier is a residual which, according to some test rule, is in contradiction to assumptions on the stochastic properties of the residuals*".

There are several methods for finding outliers such as the Z-score, modified z-score, tucked method, boxedot, MADe method, Median rule, and generalized ESD test [35]. Each test should be checked by the researcher for its suitability for the problem under consideration [36,37].

*2.2. Prediction Interval (PI)*

When attempting to forecast a value of a phenomenon, the result always has an uncertainty and an ambiguity [38], so it is necessary to define the forecasting interval.

Forecasting interval or prediction interval (PI) is defined as the interval within which the value is expected to occur with a specific probability (1–a) % [39]. It is very important to estimate this interval, in order to finally find the most accurate forecasting method [40].

The interval consists of an upper predictive limit (UPL) and a lower predictive limit (LPL) [41]. In the international literature, they are referred to either as forecast limits [42] or as prediction bounds [43].

There is a parametric and non-parametric approach for estimating the forecasting interval. In the parametric approach, which was proposed by Box and Jenkins [44], it is considered that forecasting errors $e_t = D_t - P_t$, meaning the series $\{e_t\}$, is independent and follows the normal distribution with zero mean and variance $\sigma_e^2$. The forecasting error used concerns data that is not used when creating the model or applying a method and it is often referred as an out-of-sample forecast error. The PI, for a probability (1-a) %, is calculated by the following equation [41,45]:

$$\hat{x}_N(h) \pm z_{a/2}\cdot\sqrt{\text{var}[e_N(h)]} \qquad \text{or} \qquad \hat{x}_N(h) \pm z_{a/2}\cdot\sigma_e \qquad (1)$$

where $z_{a/2}$ denotes the appropriate (two-tailed) percentage of a standard normal distribution and the standard deviation of errors is calculated from the equation:

$$\sigma_e = \sqrt{\sigma_e^2} = \sqrt{\text{var}[e_N(h)]} = \sqrt{\frac{\sum\limits_{i=1}^{N}(e_i - \hat{e})^2}{N-1}} \qquad (2)$$

and $\hat{e} = \frac{\sum\limits_{i=1}^{N} e_i}{N}$ (bibliographically referred also as bias) where N is the number of the observations.

Thus, the prediction limits are as follows: UPL $= z_{\alpha/2}\cdot\sigma_e$ and LPL $= -z_{\alpha/2}\cdot\sigma_e$.

### 2.3. The Design of the Forecasting ANN

Since the problem relates to forecasting, so that the concept time information is used, the appropriate ANN is a Dynamic Neural Network.

In this work, the dynamic network decided upon by an extensive bibliographic study is a recurrent or feedback dynamic network (RNN), because they can process data sequences, time information in our case. They can simulate two kinds of memory elements, long-term and short-term.

In every kind of ANN, the synaptic weights can be considered as long-term memories. However, if the problem has a temporal dimension, some form of short-term memory must be integrated into the neural network. The simplest way to do this is with time delays.

RNNs also allow non-linear relationships to be found, between the elements of a time series. The most important is that these networks have the advantage of recognizing time patterns regardless of their duration. More specifically; non-linear autoregressive recurrent networks (NAR) are chosen for forecasting one time step, and non-linear autoregressive with eXogenous inputs (NARX) for multi-step forecasting.

In the case of NAR, the forecasting of the $\ell(t)$ element of the timeseries is made, using only d past values of that timeseries, according to the equation:

$$\ell(t) = f(\ell(t-1), \ldots, \ell(t-d)) \tag{3}$$

Similarly, in the case of NARX, the d past values of the same timeseries ($\ell$) and also d of another exogenous timeseries ($q$) are used, according to the equation:

$$\ell(t) = f(\ell(t-1), \ldots, \ell(t-d),\ q(t-1), \ldots, q(t-d)) \tag{4}$$

The training of the ANN is done through supervised process based on prior knowledge of the right outputs. For this reason, a percentage of the historical data is retained at the end for evaluation and control of the results.

These networks produce a one-step-ahead forecasting only. In order to forecast a multi-step-ahead, a loop is added to the networks. That is, once the first forecast is made, it returns to the network as an input to produce the next forecast, and so on. Therefore, the network based on its own outputs (forecasted values), also forecasts the following future values.

In order to select the most suitable final network, after its general architecture has been decided, various tests of ANN's implementation are carried out, with the main objective of finding the one that will achieve the best results for forecasting. The basic principle used is that of validation by varying the hyper-parameters of the network: the number of hidden layers, the number of neurons of each hidden layer, the learning algorithm, the activation function of the hidden and exit neurons, and the number of delays.

- Normalization of data.

For this methodology the Min-Max normalization is proposed, in order to normalize the values between low = 0.1 and high = 0.9 using the following equation.

$$\widehat{\ell} \ = \ \text{low} \ + \ \frac{\ell - \ell_{\min}}{\ell_{\max} - \ell_{\min}} \cdot (\text{hight} - \text{low}) \tag{5}$$

where $\ell_{\min}$ and $\ell_{\max}$ are the min and the max value of the original timeseries

- Separation of training, testing and validation data sets.

The available historical data (timeseries) is proposed to be separated in chronological order by using the following percentages: training set 70%, test set 15% and validation set 15%.

- Tests on the selection of the training algorithm, the transfer function, the number of hidden neurons and the hidden levels.

The training algorithm, which is found to be appropriate, is the Bayesian regularization and the training function is the log-sigmoid (logsig) function. The use of only one hidden layer is proposed in order to keep complexity low and training times small.

*2.4. Forecasting Evaluation*

Estimating the uncertainty of the forecasts is an integral part of the forecasting process. In order to evaluate a forecast, the results obtained must be compared with the actual values, which are already known. Thus, some statistical indicators or other evaluation criteria are used.

The evaluation criteria which are proposed are the Mean Square Error (MSE), the Root of the Mean Square Error (RMSE), the Mean Absolute Error (MAE) and the Coefficient of Determination ($R^2$) [46].

$R^2$ is used as one of the usual evaluation methods is to perform a linear regression in order to check if the selected method has a good performance during the evaluation phase. This index is interpreted as the percentage of the variance of the values of the dependent variable (i.e., the predicted values) from the values of the independent variable (i.e., the actual-known values).

$R^2$ index measures the amount of variance in the target variable that can be explained by the model. In some cases, it is necessary to quantifying the error in the same measuring unit of the variable. This problem is solved by the use of RMSE. The issue with RMSE and MSE, is that since they square the residuals they tend to be more affected by large residuals. However, MSE has the possibility to highlight possible extreme deviations. This is the most common index of error that shows the qualitative performance of the forecasting method. Obviously, it does not have the same units as the actual values and the forecasts, so instead of this RMSE is used.

To solve the problem with large residuals MAE can be used, as it averages absolute value of the calculated residuals. Taking the square root of the average squared errors has some interesting implications for RMSE.

MAE is a failure index of the forecast by its actual desired value. Absolute values are used, so the forecast trend is not taken into account. The main difference between using MAE or RMSE is that MAE assigns equal weight to all errors when calculating overall performance while, as it is referred above, RMSE gives greater weight to errors with greater absolute value. Thus, RMSE is never lower than MAE for the same data set. When both metrics are calculated, RMSE is by definition never smaller than MAE. Since the errors are squared before they are averaged; RMSE gives a relatively high weight to large errors. This means RMSE should be more useful when large errors are particularly undesirable. There are cases when MAE may be steady but RMSE may increase as the variance associated with the frequency distribution of error also increases [47].

In order to define the accuracy of the forecasting the mean position change must be calculated. Specifically, a forecast can be considered as accurate if MAE is significantly smaller than the mean forecasted position change.

## 3. Case Study

GNSS data are used in numerous of studies for the monitoring of the points position in a construction or for the Earth's crustal movements [48–51]. The main advantage is that these data are in timeseries form, which is necessary for the implementation of the methodology. The above mentioned and analyzed methodology was tested by using actual data, acquired by permanent GNSS stations.

Specifically, the case study was implemented by using data from the Earthscope-Plane Boundary Observatory (PBO), which collects archives and distributes data from a broad network of over 1000 GNSS and seismic sites located across the continental United States and Alaska. These data are freely and openly available to the public [52,53] and they are referred to long time period (from 2003 to 2016).

As the methodology suggests, non-linear autoregressive recurrent networks (NAR) were used for one step forecasting, and non-linear autoregressive with exogenous inputs (NARX) either for one step or for multi-step forecasting. In order to find all the hyper parameter of the final ANN a great number of tests were carried out. All the tested ANNs were consisted of one hidden layer and the activation function of the hidden and output neurons was the logsig function. The tests were initially differentiated in the following:

- the input data normalization interval $[-0.8, 0.8]$ or $[0.1, 0.9]$,
- the input data division method, into random separation (70%, 15%, 15%) or chronological (70%, 15%, 15%),
- the training algorism trainlm or trainbfg or trainbr.

From the above combinations, 12 test sets are created. Each test set eventually implements 12 separate ANNs which then differentiate into the following parameters:

- the number of hidden neurons: min = 5, max = 20, step = 5 (four combinations)
- the number of delays: min = 2, max = 82, step = 40 (three combinations)

Additionally tests were conducted for different input and output data, according to 14 different scenarios which are created for each i GNSS station (Table 1).

**Table 1.** Scenarios of different inputs and outputs with NAR and NARX neural networks for one-step and multi-step forecasting.

| NAR | | | NARX | | | |
|---|---|---|---|---|---|---|
| Scenarios | Input Data | Output Data | Scenarios | Input Data | Exogenous Input Data | Output Data |
| I | X | X | VIII | X | $\sigma_X$ (station i) | X |
| II | Y | Y | IX | Y | $\sigma_Y$ (station i) | Y |
| III | Z | Z | X | Z | $\sigma_Z$ (station i) | Z |
| IV | X, Y, Z | X, Y, Z | XI | X | Mean: dX (neighbor stations) | X |
| V | X, $\sigma_X$ | X, $\sigma_X$ | XII | Y | Mean: dY (neighbor stations) | Y |
| VI | Y, $\sigma_Y$ | Y, $\sigma_Y$ | XIII | Z | Mean: dZ (neighbor stations) | Z |
| VII | Z, $\sigma_Z$ | Z, $\sigma_Z$ | XIV | X, Y, Z | Mean: dX, dY, dZ (neighbor stations) | X, Y, Z |

Therefore, 12 test sets were created for each of the 14 scenarios. In each test set, 12 ANNs with different parameters are implemented and trained for each GNSS station. In total 2016 (12 test sets × 14 scenarios × 12 ANNs) ANNs were tested only for one step forecasting for each GNSS permanent station. Furthermore, the same number of ANNs were trained for multi-step forecasting, with the difference that in this case only NARX neural networks where used.

Figure 2 presents examples of the mentioned in paragraph 2.4 evaluation criteria for three random GNSS stations, namely AC57, P515 and P788. The final hyper-parameters are chosen after the assessment of those diagrams. Finally, the network, which has the smallest values (for the test set) in these criteria, is selected. Each GNSS station has different number of test set data, because this number depends from the original available data. However, in every case the test set is the 70% of the total available data. In particular, the exact number of the training set was ranged between 300 daily records and 2000 depending each GNSS station.
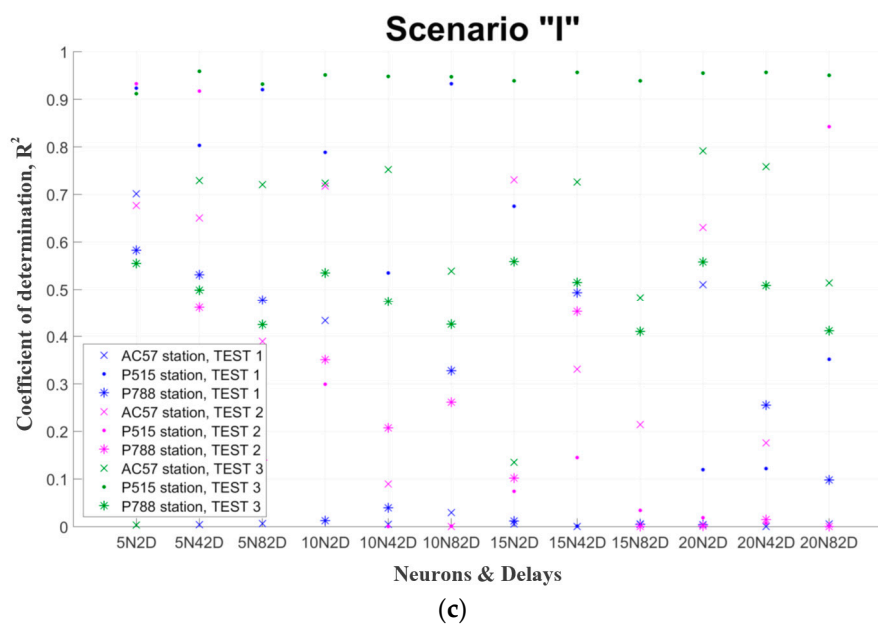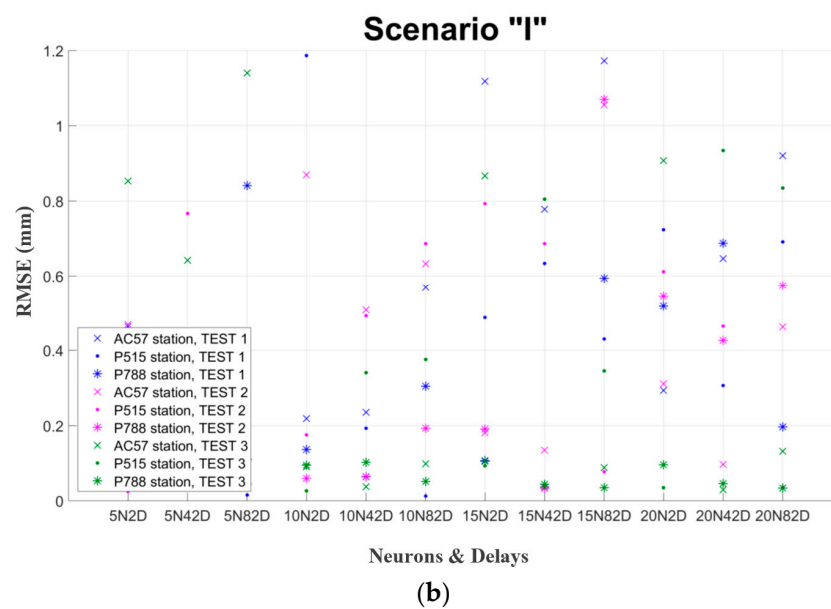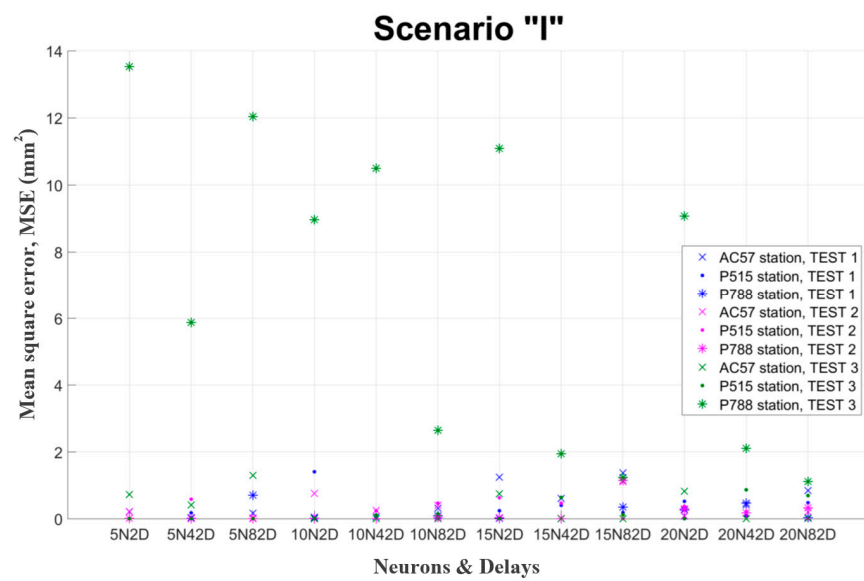
(a)



(b)



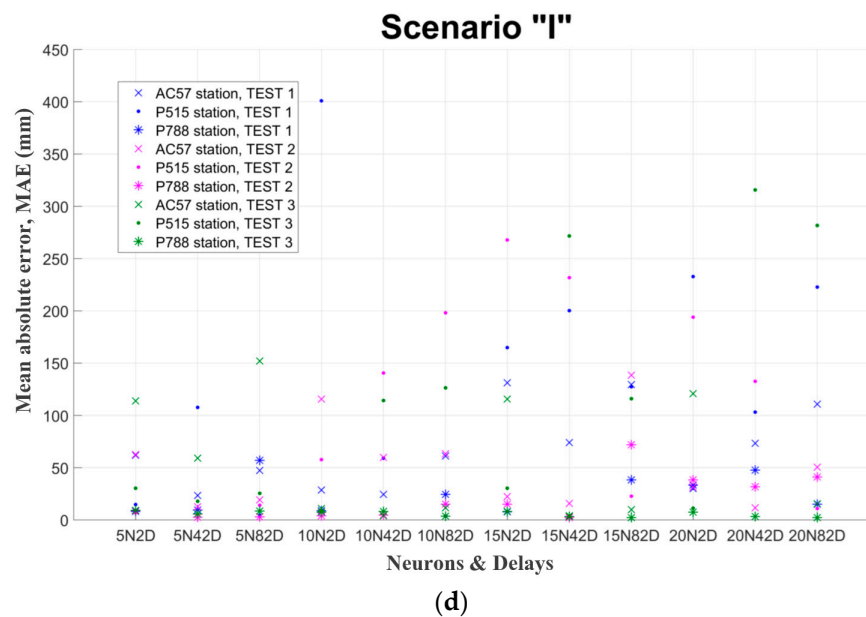(c)

**Figure 2.** *Cont.*

**(d)**

**Figure 2.** Combined diagrams (**a**) MSE; (**b**) RMSE; (**c**) $R^2$; (**d**) MAE.

After all the above evaluation it was resulted that the best ANN for point position forecast: has as inputs and outputs separately the geocentric coordinates X, Y, Z (m) of each station (scenarios I, II and III). It consists of one hidden layer and has as activation function of hidden neurons and output neurons the logsig function. Data is separated in chronological order (70%, 15%, and 15%) and their normalization is in the range [0.1,0.9]. The training algorithm is Bayesian regularization. The number of hidden neurons is 10 and the number of delays 42 for one-step forecasts. While for multi-step forecasts there are 20 neurons and 2 delays.

The mean MAE of the 1000 GNSS permanent stations was calculated and presented in the following table (Table 2).

**Table 2.** Mean, max and min MAE of the proposed best ANN for short-term and long-term forecasting and of the position change dX, dY and dZ for the 1000 GNSS permanent stations.

| Coordinate | MAE (mm) One-Step | | | MAE (mm) Multi-Step | | | Position Change (mm) | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | **Mean** | **Max** | **Min** | **Mean** | **Max** | **Min** | **Mean** | **Max** | **Min** |
| dX (I) | 0.5 | 1.5 | 0.2 | 1.2 | 2.0 | 0.5 | 1.8 | 11.6 | 0.7 |
| dY (II) | 0.9 | 1.1 | 0.3 | 1.5 | 1.8 | 0.4 | 2.6 | 7.5 | 0.8 |
| dZ(III) | 1.0 | 1.6 | 0.5 | 1.8 | 1.8 | 0.6 | 2.6 | 12.4 | 1.7 |

In addition some other conventional methods were used for the same purpose. The scope is to compare the forecasts generated by this intelligent method to those, which are provided by some other specific conventional methods. The nine conventional methods that were selected to be examined and compared are the following: Kinematic linear model, Kinematic polynomial 2nd degree model, Simple mean, Simple moving average, Simple exponential smoothing, Double moving average, Double exponential smoothing (Brown's method), Exponential smoothing adjusted for trend (Holt's method) and Autoregressive Moving Average Methods (ARMA). The comparative results of the chosen conventional methods and the proposed ANN are presented in the following Figure 3.

Additionally Figure 4 illustrates the reliability of the ANNs forecasts relative to the two optimum, between the nine chosen ones, conventional methods (kinematic linear model, ARMA (4,3). It is obvious that the proposed ANN, especially in the case of short-term forecasts, is able to adequately forecast the size and the trend of the change of point positions. On the contrary, it is also evident that the two examined conventional methods have a significant deviation both in the forecast of the size and of the trend of the point positions change.
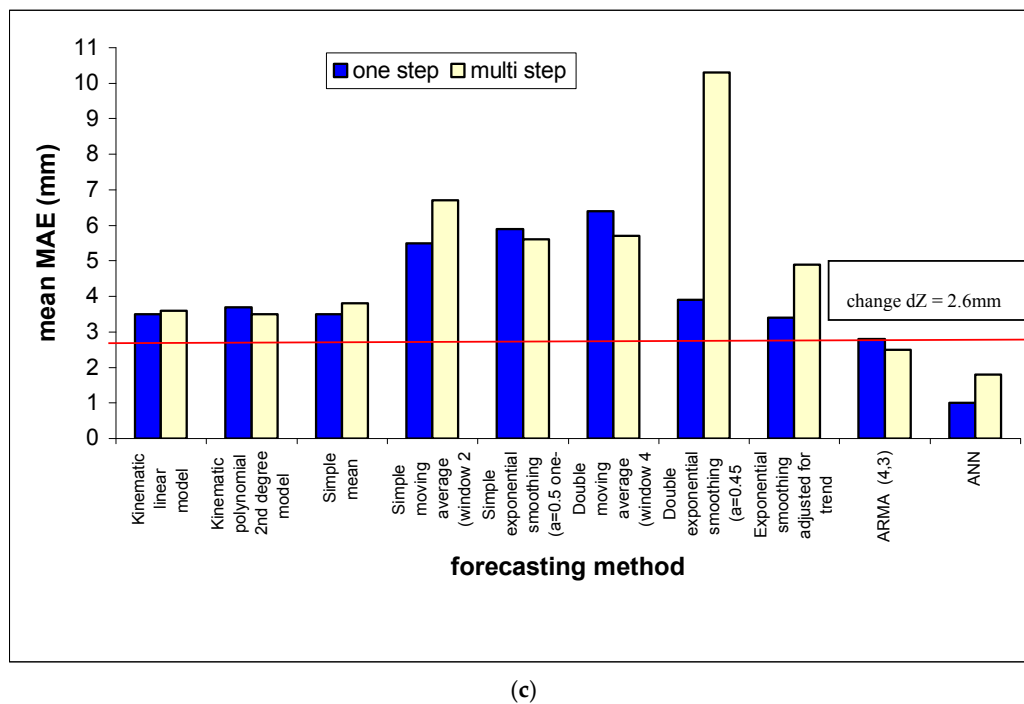


(**a**)



(**b**)

**Figure 3.** *Cont.*

(**c**)
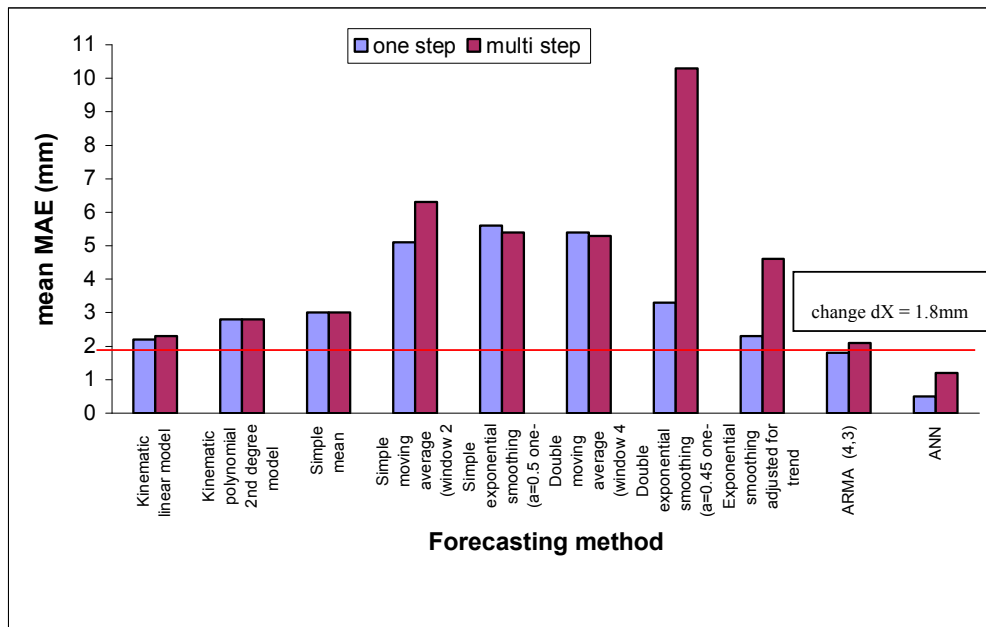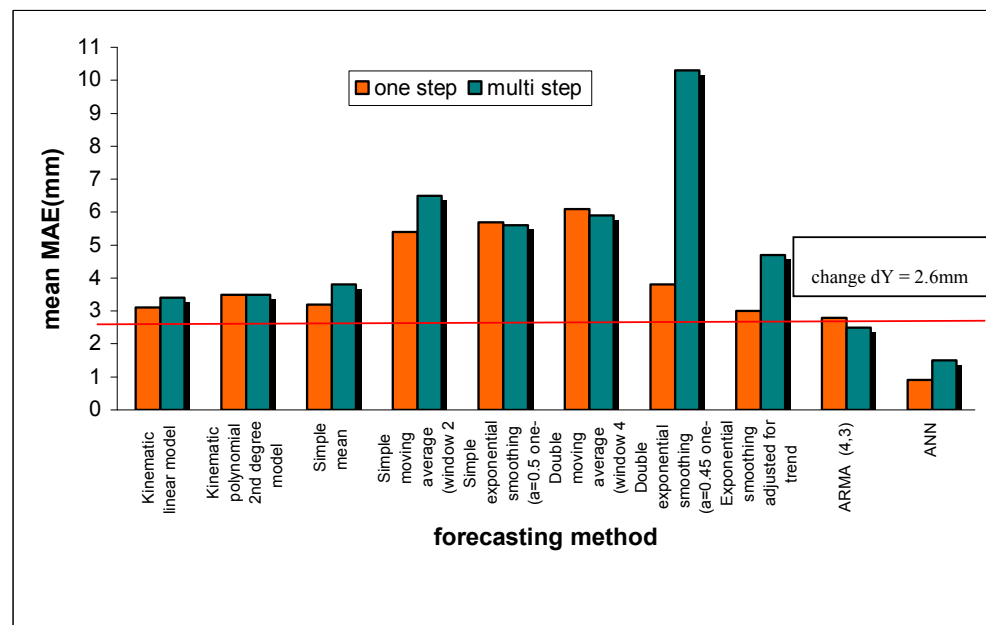
**Figure 3.** Mean MAE for the proposed ANN and the examined conventional methods for dX (**a**), dY (**b**) and dZ (**c**).
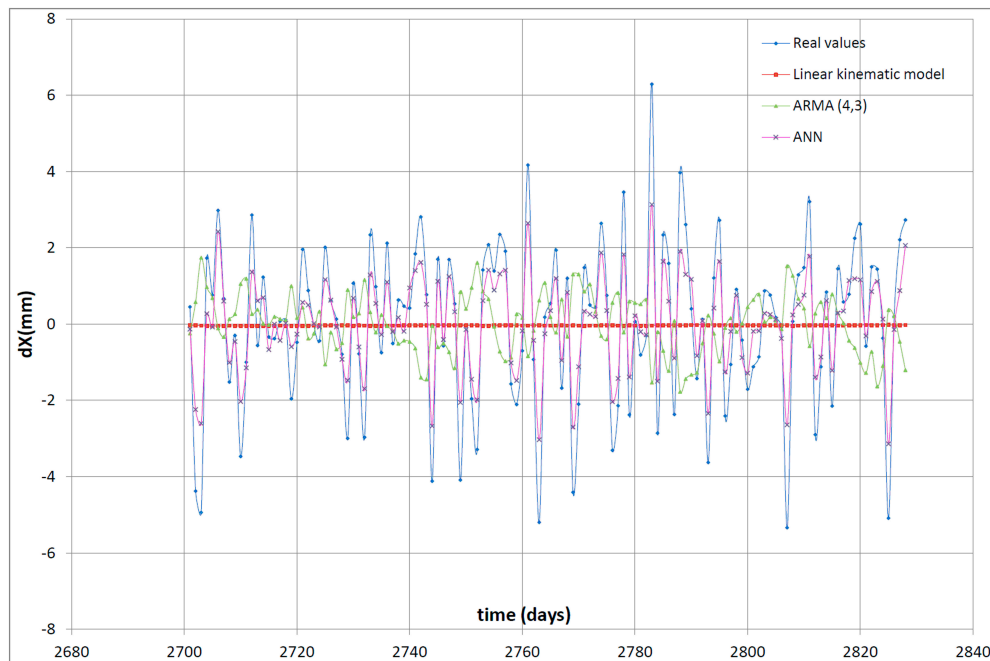


**Figure 4.** Time series of short-term forecasts of the linear kinematic model, the ARMA (4,3) and the proposed ANN for dX (mm).

## 4. Discussion

The proposed methodology has been developed to solve the problem both for one-step-ahead (short-term) and for multi-step-ahead (long-term) forecasting.

To sum up, the ANN resulting as the best for forecasting the change of points position is a non-linear autoregressive NAR recurrent neural network and a NARX non-linear autoregressive with

eXogenous inputs in the case of multi-step forecasting. Some preliminary tests on a sample dataset led to the following ANN design choices and training procedure:

- It consists of one hidden layer and has "logsig" as an activation function of hidden and output neurons.
- Separation of the data is done in chronological order (70%, 15%, 15%) and their normalization in the interval [0.1,0.9]
- The learning algorithm is the Bayesian regularization backpropagation.
- It is proposed the number of hidden neurons to be 10 and the number of delays 42 for one-step forecasts, while for multi-step forecasts 20 hidden neurons and 2 delays.
- Finally, in cases where time series of geocentric coordinates X, Y, Z (m) are used, it is proposed X, Y, Z to be used separately as inputs and outputs.

The best ANNs for one-step and for multi-step forecasting are presented accordingly in the following Figure 5.
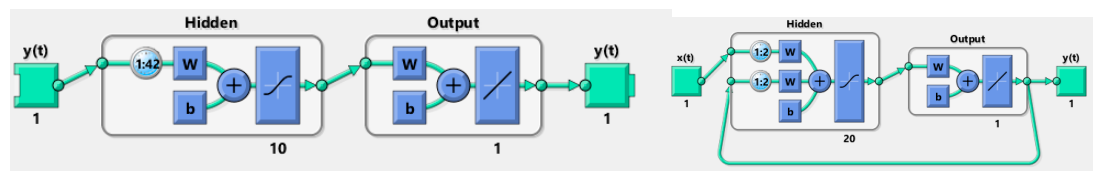


**Figure 5.** The NAR for one-step and NARX for multi-step forecasting suggested networks for the forecasting of points position changing.

The number of available data to be used is a very important factor in their training. As the number of data increases, the better the network will be "trained" and will therefore produce better results.

Big data management is very complex and requires special treatment. In such cases it is necessary to use automated techniques rather than optical means as in smaller volume and size data. Despite the difficulties, finding and using large amounts of information is a major asset especially when it comes to forecasting.

During training, special care must be taken to avoid the overfitting phenomenon. If the network is trained for too many epochs and/or in the entire dataset, we could observe small training error but lose the ability to generalize well, i.e., perform well on new, unseen data. Therefore, there is a possibility that there is a limit beyond which, while the training error decreases, the validation error increases. In this situation, the network may have memorized the correct results that arise from the education phase, but has not learned to generalize properly on new data.

As already mentioned, the proposed methodology is based on ANNs. The main drawback of their use is the large number of tests (try-and-error process) that must be carried out with all the possible different combinations of the hyper-parameters.

Therefore, in cases where the data is so many, a representative sample should be selected by which all the experiments will be carried out and, finally, the most suitable ANN will be proposed. The sample could be simple random, systematic or else quasi-random and finally stratified random depending on the available dataset.

Despite this disadvantage, the methodology gives more accurate forecasts compared to some other selected conventional-statistical forecasting methods (i.e., simple mean, simple moving average, simple exponential smoothing, double moving average, Brown's method, Holt's method, ARMA/ARIMA models) [54].

## 5. Conclusions

ANNs have been introduced to the engineering sciences in general and Geodesy in particular. In recent years, an attempt has been made to investigate their use. Several of the geodetic applications

in which the ANNs have been used to date have yielded very satisfactory results, giving even better results compared to classical methods for solving these problems. Applications mainly used are local geoid mapping, averaging of the sea level, coordinate conversion to reference systems, mobile mapping models, and so on.

This research, by using actual data, acquired by 1000 permanent GNSS stations, ascertains that the best forecasting results succeeded when simple scenario of input and output data are applied, namely when the forecasting is made separately for each coordinate X or Y or Z. This is concluded after 14 different scenarios results. Also the data separation in chronological order leads to better forecasts than the random separation.

The numerous available data permits to realize all the 12 test sets, which are created for each of the 14 scenarios and the 12 ANNs with different parameters in each test set, in total 2016 ANN. This gives the opportunity for reliable conclusions. MAE is proved to be a satisfied and unbiased index as assigns equal weight to all errors when calculating overall performance, so it provides a real estimation of the forecast uncertainty. The multi step forecast provides always twice greater MAE than the one step ahead.

The ANN can forecast a point's position change of the order of 2 mm with MAE in the order of 0.5 mm, while the optimum convention method, ARMA model (4,3), with MAE 1.8 mm. Moreover the most significant is that it gives not only the right size of the position change with adequate accuracy but also it indicates the right trend of the position change vector.

As far as ANNs are concerned, this is a modern and very attractive intelligent technique. For this reason, implementing an ANN to solve any problem is an interesting and original alternative. However, a try-and-error process must always be followed during the implementation process. The more testing done, the more complete the work is to investigate the use of ANNs for any problem. This means that it takes a lot of time and a lot of computational power until one gets the desired result.

The documentation and suggestion of a particular ANN and all its hyper-parameters arises after each application, which is another significant disadvantage. ANNs are a data driven method, so the architecture and hyper-parameters found to be optimal in a dataset, in the general case won't be optimal for another dataset.

This methodology, which is based on ANNs is accurate, stable and general and can be used in order to forecast the changing of points' position. It also gives accurate and reliable results compared to some selected classical-conventional methods.

## References

1. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [CrossRef]
2. Yang, C.-C.; Prasher, S.O.; Landry, J.-A.; Ramaswamy, H.S.; Ditommaso, A. Application of artificial neural networks in image recognition and classification of crop and weeds. *Can. Agric. Eng.* **2000**, *42*, 147–152.
3. Babu, B.V.; Ramakrishna, V.; Kalyan Chakravarthy, K. Artificial Neural Networks for Modeling of Adsorption. In Proceedings of the Second International Conference on Computational Intelligence, Robotics, and Autonomous Systems (CIRAS-2003), Singapore, 15–18 December 2003.
4. Khalil, B.M.; Awadallah, A.G.; Karaman, H.; El-Sayed, A. Application of Artificial Neural Networks for the Prediction of Water Quality Variables in the Nile Delta. *J. Water Resour. Prot.* **2012**, *4*, 388–394. [CrossRef]
5. Memarianfard, M.; Hatami, A.M. Artificial neural network forecast application for fine particulate matter concentration using meteorological data. *Glob. J. Environ. Sci. Manag.* **2017**, *3*, 333–340. [CrossRef]
6. Khaze, S.R.; Masdari, M.; Hojjatkhah, S. Application of Artificial Neural Networks in estimating participation in elections. *Int. J. Inf. Technol. Model. Comput.* **2013**, *1*, 23–31. [CrossRef]

7.    Amato, F.; López, A.; Peña-Méndez, E.; Vaňhara, P.; Hampl, A.; Havel, J. Artificial neural networks in medical diagnosis. *J. Appl. Biomed.* **2013**, *11*, 47–58. [CrossRef]

8.    Veronez, M.R.; de Souza, S.F.; Matsuoka, M.T.; Reinhardt, A.; de Silva, R.M. Regional Mapping of the Geoid Using GNSS (GPS). Measurements and an Artificial Neural Network. *Remote Sens.* **2011**, *3*, 668–683. [CrossRef]

9.    Makarynskyy, O.; Makarynska, D.; Kuhn, M.; Featherstone, W.E. Predicting sea level variations with artificial neural networks at Hillarys Boat Harbour Western Australia. *Estuar. Coast. Shelf Sci.* **2004**, *61*, 351–360. [CrossRef]

10.    Turgut, B. A back-propagation artificial neural network approach for three-dimensional coordinate transformation. *Sci. Res. Essays* **2010**, *5*, 3330–3335.

11.    Miczulski, W.; Sobolewski, L. Influence of the GMDH Neural Network data preparation method on UTC (PL) correction prediction results. *Metrol. Meas. Syst.* **2012**, *19*, 123–132. [CrossRef]

12.    Luzar, M.; Sobolewski, L.; Miczulski, W.; Korbicz, J. Prediction of corrections for the Polish time scale UTC(PL) using artificial neural networks. *Bull. Pol. Acad. Sci. Tech. Sci.* **2012**, *61*. [CrossRef]

13.    Minsky, M.L. Steps towards artificial intelligence. *Proc. Inst. Radio Eng.* **1961**, *49*, 8–30. [CrossRef]

14.    Patterson, D.W. *Artificial Neural Networks: Theory and Applications*; Prentice Hall: Upper Saddle River, NJ, USA, 1996.

15.    Haykin, S. *Neural Networks: A Comprehensive Foundation*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999.

16.    Basheer, I.A.; Hajmeer, M. Artificial neural networks: Fundamentals, computing, design, and application. *J. Microbiol. Methods* **2000**, *43*, 3–31. [CrossRef]

17.    Hudson, D.; Cohen, M. *Neural Networks and Artificial Intelligence for Biomedical Engineering*; The Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2000.

18.    Kriesel, D. A brief Introduction to Neural Network. 2005. Available online: http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf (accessed on 9 March 2018).

19.    Haykin, S. *Neural Networks and Learning Machines*, 3rd ed.; Pearson Education: London, UK, 2010; Available online: http://www.dkriesel.com/en/science/neural_networks (accessed on 8 March 2018).

20.    Armstrong, J.S.; Green, C.K.; Graefe, A. *Forecasting Principles*; Scholarly Commons; University of Pennsylvania: Philadelphia, PA, USA, 2010.

21.    Fildes, R. *Bibliography on Business and Economic Forecasting*; Working Paper; Lancaster Center for Forecasting, Lancaster University Management School: Lancaster, UK, 2014.

22.    Ghorbani, E.; Cha, Y.J. An iterated cubature unscented Kalman filter for large-DoF systems identification with noisy data. *J. Sound Vib.* **2018**, *420*, 21–34. [CrossRef]

23.    Cha, Y.J.; Chen, J.G.; Büyüköztürk, O. Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters. *Eng. Struct.* **2017**, *132*, 300–313. [CrossRef]

24.    Lewis-Beck, M. Election of Forecasting: Principles and Practice. *Br. J. Politics Int. Relat.* **2005**, *7*, 145–164. [CrossRef]

25.    Frawley, W.; Piatetsky-Shapiro, G.; Matheus, C. Knowledge Discovery in Databases: An Overview. *AI Mag.* **1992**, *13*, 57–70.

26.    Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. From Data Mining to Knowledge Discovery in Databases. *AI Mag.* **1996**, *17*, 37–54.

27.    Bellazzi, R.; Zupan, B. Predictive data mining in clinical medicine: Current issues and guidelines. *Int. J. Med. Inform.* **2008**, *77*, 81–97. [CrossRef] [PubMed]

28.    Chakravarti, I.M.; Laha, R.G.; Roy, J. *Handbook of Methods of Applied Statistics*; John Wiley and Sons: Hoboken, NJ, USA, 1967; Volume I, pp. 392–394.

29.    Stephens, M.A. EDF Statistics for Goodness of Fit and Some Comparisons. *J. Am. Stat. Assoc.* **1974**, *69*, 730–737. [CrossRef]

30.    Box, G.E.P.; Cox, D.R. An analysis of transformations. *J. R. Stat. Soc.* **1964**, *26*, 211–234.

31.    Osborne, J.W. Improving your data transformations: Applying the Box-Cox transformation. *Pract. Assess. Res. Eval.* **2010**, *15*, 1–9.

32.    Kovarik, M.; Sarga, L. Process capability indices for non-normal data. *WSEAS Trans. Bus. Econ.* **2014**, *11*, 419–429.

33.    Chatfield, C. *The Analysis of Timeseries an Introduction*, 6th ed.; Chapman Hall/CRC: London, UK, 2003.

34.  Caspary, W. *Concepts of Networks and Deformation Analysis*; Monograph No. 11; School of Surveying, University of New South Wales: Sydney, Australia, 1987.

35.  Rosner, B. Percentage Points for a Generalized ESD Many-Outlier Procedure. *Technometrics* **1983**, *25*, 165–172. [CrossRef]

36.  Fildes, R.; Hibon, M.; Makridakis, S.; Meade, N. Generalizing about univariate forecasting methods: Further empirical evidence. *Int. J. Forecast.* **1998**, *14*, 339–358. [CrossRef]

37.  Wdowinski, S.; Bock, Y.; Zhang, J.; Fang, P.; Genrich, J. Southern California permanent GPS geodetic array: Spatial filtering of daily position for estimating coseismic and postseismic displacements induced by the 1992 Landers earthquake. *J. Geophys. Res.* **1997**, *102*, 18057–18070. [CrossRef]

38.  Pinson, P. Estimation of the Uncertainty in Wind Power Forecasting. Ph.D. Thesis, Ecole des Mines de Paris, Center for Energy and Processes, Rue Claude Daunesse, Sophia Antipolis, Paris, France, 2006.

39.  Pinson, P.; Kariniotakis, G. On-line adaptation of confidence intervals based on weather stability for wind power forecasting. In Proceedings of the 2004 Global Windpower Conference, Chicago, IL, USA, 28–31 March 2004.

40.  Holfmann-Wellenhof, B.; Moritz, W.E. *Physical Geodesy*, 2nd ed.; Springer Wien: New York, NY, USA, 2006.

41.  Chatfield, C. Prediction intervals for time-series forecasting. Department of Mathematical Science. University of Bath. In *Principle of Forecasting: A Handbook for Researchers and Practitioners*; Armstrong, S.J., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001.

42.  Wei, W.W.S. *Timeseries Analysis*; Addison-Wesley: Redwood City, CA, USA, 1990.

43.  Brockwell, P.J.; Davis, R.A. *Timeseries: Theory and Methods*, 2nd ed.; Springer-Verlag: New York, NY, USA, 1991.

44.  Box, G.E.P.; Pierce, D.A. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *J. Am. Stat. Assoc.* **1970**, *65*, 1509–1526. [CrossRef]

45.  Turyna, M.; Hrdina, T. *Calculating Interval Forecasts*; Department of Economics, University of Vienna: Vienna, Austria, 2009.

46.  Alevizakou, E.G.; Pantazis, G. A Comparative Evaluation of Various Models for Prediction of Displacements. *Appl. Geomat.* **2017**, *9*, 93–103. [CrossRef]

47.  Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]

48.  Bell, J.W.; Amelung, F.; Ramelli, A.R.; Blewitt, G. Land subsidence in Las Vegas, Nevada, 1935–2000: New geodetic data show evolution, revised spatial patterns and reduced rates. *Environ. Eng. Geosci.* **2002**, *8*, 155–174. [CrossRef]

49.  Chan, W.S.; Xu, Y.L.; Ding, X.L.; Xiong, Y.L.; Dai, W.J. Dynamic displacement measurement accuracy of GPS for monitoring large civil engineering structures. In Proceedings of the SPIE, International Symposium on Smart Structures and Nondestructive Evaluation (CDROM), San Diego, CA, USA, 17 May 2005; SPIE: Bellingham, WA, USA, 2005.

50.  Psimoulis, P.; Ghilardi, M.; Fouache, E.; Stiros, S. Subsidence and evolution of the Thessaloniki plain, Greece, based on historical leveling and GPS data. *Eng. Geol.* **2007**, *90*, 55–70. [CrossRef]

51.  Farmer, G.T.; Newman, A.V.; Psimoulis, P.; Stiros, S. Geodetic characterization of Santorini Caldera from GPS measurements. Abstract G43B-1195. In Proceedings of the AGU-2007 Fall Meeting, San Francisco, CA, USA, 10–14 December 2007.

52.  EarthScope—Data. Available online: http://www.earthscope.org/node/395 (accessed on 8 March 2018).

53.  EarthScope—Plate Boundary Observatory (PBO). Available online: http://www.unavco.org/projects/major-projects/pbo/pbo.html (accessed on 8 March 2018).

54.  Alevizakou, E.G. Development of a Methodology Using Artificial Neural Networks to Forecast the Change of Point Positions in Space. Ph.D. Thesis, NTUA, Athens, Greece, 2017.