FORECASTING METHODS FOR SPARE PARTS DEMAND

Relatore: Ch.mo Prof. Maurizio Faccio
Correlatore: Dott. Ing. Fabio Sgarbossa
Laureando: Andrea Callegaro
Matricola: 580457

ANNO ACCADEMICO 2009/2010
INDEX

Summary........................................................................................................................................4
Introduction......................................................................................................................................5
Chapter 1 - Analysis of spare parts and spare parts demand................................. 7
1. Introduction.................................................................................................................................. 7
2. Spare parts features..................................................................................................................... 8
3. Spare parts demand and classifications................................................................................... 9
    3.1. Analytic Hierarchy Process................................................................................................. 12
4. Annexed costs............................................................................................................................. 13
Chapter 2 - An overview of literature on spare parts demand forecasting
    methods........................................................................................................................................... 15
1. Introduction................................................................................................................................ 15
2. Forecasting methods in the scientific literature................................................................. 15
3. Explanation of forecasting methods...................................................................................... 23
    3.1. Single exponential smoothing ........................................................................................... 23
    3.2. Croston’s method................................................................................................................ 23
    3.3. Syntetos – Boylan Approximation ..................................................................................... 24
    3.4. Moving Average .................................................................................................................. 25
    3.5. Weighted moving average .................................................................................................. 26
    3.6. Additive and multiplicative winter .................................................................................... 26
    3.7. Bootstrap method ............................................................................................................... 27
    3.8. Poisson method ................................................................................................................. 28
    3.9. Binomial method ............................................................................................................... 29
    3.10. Grey prediction model ...................................................................................................... 30
    3.11. ARMA(p,q) ARIMA(p,d,q) S-ARIMA(p,d,q)(P,D,Q)s .................................................. 31
    3.12. Neural networks .............................................................................................................. 33
4. Benchmarks.................................................................................................................................. 34
    4.1. MAPE ................................................................................................................................. 34
    4.2. S-MAPE ............................................................................................................................ 35
    4.3. A-MAPE ........................................................................................................................... 35
    4.4. RMSD ............................................................................................................................... 36
    4.5. RGRMSE .......................................................................................................................... 36
    4.6. PB ..................................................................................................................................... 36
Chapter 3 - Neural networks in spare parts forecasting....................................... 38
1. Introduction................................................................................................................................. 38
2. What are neural networks?........................................................................................................ 38
2. Benefits of neural networks...................................................................................................... 40
SUMMARY

The following work of thesis deals with the methods of forecasting the future spare parts demand and, in particular the application of neural networks in this field. Before examining the methods of forecast, a first chapter is dedicated to the presentation of the problem and the individuation of the spare parts and spare parts demand features. The first chapter gives a panoramic on spare parts and the way in which they get into the field of industrial maintenance; in particular first chapter deals with spare parts features, analysis of spare parts demand, classifications and annexed problems, as annexed costs and management policies.

The central part, after a brief introduction on the different kinds of spare parts management, gives an overview of recent literature in the field of spare parts demand forecasting methods: all the most studied methods are treated and, for everyone of them, among other things, a brief explanation and a description of limits and innovative features are given.

Third chapter deals with neural networks and their application in the field of spare parts management: a detailed description of the background theory and a detailed explanation of the way in which they are used in this field (also through and two cases of study) are given.

Finally, real data of spare parts consumption in a business of the iron and steel sector are used to show how some of these forecasting methods may be applied in the industrial reality.
INTRODUCTION

Production and manufacturing problems received tremendous interest from the operations research and management science researchers. Many books and textbooks have been written and several journals are dedicated to these subject. These topics are part of the curriculum in various industrial, mechanical, manufacturing, or management programs.

In the past, maintenance problems received little attention and research in this area did not have much impact. Today, this is changing because of the increasing importance of the role of maintenance in the new industrial environment. Maintenance, if optimized, can be used as a key factor in organizations efficiency and effectiveness. It also enhances the ability of the organization to be competitive and meets its stated objectives.

The research in the area of maintenance management and engineering is on the rise. Over the past few decades, there has been tremendous interest and a great deal of research in the area of maintenance modeling and optimization. Models have been developed for a wide variety of maintenance problems. Although the subject of maintenance modeling is a late developer compared to other area like production systems, the interest in this area is growing at an unprecedented rate.

In particular, the availability of spare parts and material is critical for maintenance systems. Insufficient stocks can lead to extended equipment down time. On the contrary, excessive stocks lead to large inventory costs and increase system operating costs. An optimal level of spare parts material is necessary to keep equipment operating profitably.

In order to appreciate the potential of the problem of spare parts and material provisioning in maintenance and realize the benefits of optimization models in this area, Sherbrooke (1968, p.125-130) estimated that in 1968 the cost of recoverable spare parts in the United State Air Force (USAF) amounted to ten billion dollars. This cost were about 52% of the total cost of inventory for that year.

With the expansion of high technology equipment in industries world-wide the need for spare parts to maximize the utilization of this equipment is paramount. Sound spare parts management improves productivity by reducing idle machine time and increasing resource utilization. It is obvious that spares provisioning is a complex problem and requires an accurate analysis of all conditions and factors that affect the selection of appropriate spare provisioning models. In the recent literature there are large numbers of papers in the general area of spare provisioning. Most of these papers deal with the policies to assure the optimal stock level, and only in the very
last years the research began to focus on spare parts demand forecasting. In the industrial field, forecasts have a crucial role: modern organizations have to know the trend of the demand in order to plan, organize the production, and manage the stock. In the last years another aspect has emerged: the importance of increasing the value of the management aspect of the maintenance processes and services, contemplating technical, economic and organizational issues of this function; in this point of view, forecasts tied to the future utilize of spare parts get into.

But there are difficulties in forecasting the demand of spare parts; this because in process industries the characteristics and requirements for inventory management of spare parts differ from those of other materials.

Therefore, this work has the following objectives:

- to show the problem of spare parts management in the modern industries, explaining spare parts and spare parts demand features;
- to discuss and explain the spare parts demand forecasting methods that have been more studied than others in the recent scientific literature;
- to focalize into neural network models, one of the methods that have gained the most part of the scientific attention in the very last years of research;
- to show the application of these methods on real industrial data.
CHAPTER 1
Analysis of spare parts and spare parts demand

1. Introduction
In the normal life cycle of an industrial system, or simply of an equipment, as consequence of a breakdown, derived from the inevitable phenomenon of the usury, there is the necessity to replace parts or components.
For this reason the crucial problem of spare parts management fall within the maintenance problematic. Sometimes in the industrial reality this aspect is ignored but, as we’ll see, it has a great relevance both in technical and in economic point of view.
Let’s think, for example, to the features of a possible breakdown maintenance intervention (Fig. 1.1) which requests the use of consumer materials or the substitution of a part of the system.
Up-time is the time of functioning, while down-time is the time required to repair the system.

Figure 1.1 – Example of an intervention after a break-down

Within the framework of the activities executed on the occasion of a maintenance intervention there is frequently (almost always) a phase of supplying of spare parts. The duration of this phase is substantially influenced by the presence or not of the spare materials in the local warehouse of the business.
The supplying lead–time can last few minutes, if the necessary materials are in hand of the firm, some days or even some weeks in case that the business has to require an
item available in a supplier geographically very far or even not available in the supplier firm. Therefore, a burden tied to the lack of production can be associated to the last of the cycle of supplying of a spare part and, cause of the complexity achieved by the production systems, these costs might be significant.

Sometimes the presence of high down-times due to lack of spare parts is gotten around with further pejorative behaviours: for example, components similar to the originals but not adapt are assembled with the result of damaging the productive system and compromising the situation.

On the other hand, spare parts have proper features that lead them to have not certain employ on the machinery; this can be translate into high risks of obsolescence generally associated with great costs of purchasing.

In this capitol the following themes are discussed:
- the proper features of the spare parts;
- the possible classifications of spare parts;
- the costs annexed to spare parts management;
- the features of the spare parts demand.

2. Spare parts features

Spare material has peculiar characteristics that distinguish it from all the other materials used in a productive or service system. The principal feature resides in the consumption profile: the demand of spare parts (as explained after) is in the major part of the cases intermittent (an intermittent demand is a demand which takes place with irregular time intervals and concerns reduced and, above all, very variable quantities).

Another distinctive characteristic of maintenance spare parts is the specificity of the employ. Usually, spare parts aren’t of the type “general purpose” and so they have to be employed only for the use and the function for that they have been realized. This, inevitably, hide great risk of obsolescence which is experimented when the substitution of an equipment is decided: the set of spare parts that aren’t re-usable on other systems (generally, the major part) becomes immediately obsolete. In the best of the hypothesis the set might be sold contextually with the system (if this is sold) or with the articles.

The spare parts have generally a great technical content and, for this, an high unit value. Therefore they require significant financial efforts for their purchase and they consume significant costs also for their maintenance.

It is necessary to add that, often, for the storage of technical material for the spare parts, costly devices are essential; for example devices tied to the protection, or to the necessity of setting with particular conditions, or even other.
In conclusion, spare parts have particular features which make extremely delicate and sophisticated their management.

3. Spare parts demand and classifications

The spare parts demand is very particular. In the majority of the cases, it takes place with irregular time intervals and concerns reduced and, above all, very variable quantities, as shown in Figure 1.2.

\[ \varepsilon_i = \text{consumption of spare part (pieces)} \]
\[ t_i = \text{interval between two consecutive demands} \]

*Figure 1.2 – example of the intermittent consumption of a spare part*

For a valuation of this double characterization of spare parts demand, two parameters recognized in international field are utilized:

- **ADI** - *Average inter-demand interval*: average interval between two demand of the spare part. It is usually expressed in periods, where the period is the referential time interval which the business utilizes for the purchases;
- **CV** – *Coefficient of variation*: standard deviation of the demand divided by the average demand.

\[
 \text{ADI} = \frac{\sum_{i=1}^{N} t_i}{N} \tag{1.1}
\]
\[
 \text{CV} = \sqrt{\frac{\sum_{i=1}^{N} (\varepsilon_i - \bar{\varepsilon})^2}{\frac{N}{\varepsilon}}} \tag{1.2}
\]
\[
 \text{where } \varepsilon = \frac{\sum_{i=1}^{N} \varepsilon_i}{N} \tag{1.3}
\]
For ADI, N is the number of periods with non-zero demand, while for CV it is the number of all periods.

Ghobbar et al. (2003, p.2105) suggest some “cut values” which allow a more detailed characterization of the intermittent standard of spare parts demand. The Figure 1.3 presents the four categories of the spare parts demand (patterns) as they are defined by the present literature:

Four typologies can be recognized:

- **Slow moving (or smooth)**: this items have a behaviour which is similar to that of the traditional articles, at low rotation, of a productive system;
- **Strictly intermittent**: they are characterized by extremely sporadic demand (therefore a lot of a period with no demand) with a not accentuated variability in the quantity of the single demand;
- **Erratic**: the fundamental characteristic is the great variability of the requested quantity, but the demand is approximately constant as distribution in the time;
- **Lumpy**: it is the most difficult to control category, because it is characterized by a lot of intervals with zero-demand and a great variability in the quantity.

This first subdivision is functional to the research of the different forecasting methods for the different categories so to obtain the best possible performances in the difficult process of the analysis and estimate of the requirements.

It is important to say that this categorization was after criticized by Syntetos (2007, p.166-167); he asserted that this scheme was developed under the premise that is preferable to identify conditions under which one method outperforms one or more other estimators and then categorize demand based on the results, rather than working the other was around, as it is often happening in practice. Both the parameters and their cut-off values were the outcome of a formal mathematical analysis of the mean-squared error associated with three intermittent demand estimators: single exponential
smoothing (SES), Croston’s method and SBA (they are explained in chapter 2). He also asserted that Ghobbar tried to identify useful categorisation parameters rather than the specification of their exact cut-off values, that may differ from one situation to the other (especially if different estimators are considered).

However, there are other important factors such as cost and criticality of the part, which influence the decisions to take in the field of spare parts management (for example: “how much to order”, “when to order”, …). Therefore, spare parts can also be classified in terms of cost and criticality.

The cost relates to purchase and maintenance cost and can be classified as low, moderate or high (Ben-Daya et al. 2000, p.99).

Criticality is based on the risk (and cost) of not completing the process or assigned equipment function i.e. “the mission”. Also criticality can be classified as low, moderate or high.

Highly critical spare parts are those which are absolutely essential for mission success. Moderately critical parts are such that if they are out of stock at the time of demand, it will have only a slight to moderate effect on mission success, whereas parts of low criticality are not absolutely essential for mission success. If such parts of low criticality aren’t available on demand, alternate parts can be substituted, or in-plant manufacturing of such parts is possible, or they are instantly available in the market. To assess the criticality of spare parts there are several methods: one may be the Pareto analysis of the part/equipment to establish ABC classification, and development of a criticality or cost/criticality loss matrix. The Fig.1.4 is an example of cost criticality/loss matrix proposed by Ben-Daya et al. 2000, p.100.

<table>
<thead>
<tr>
<th>CRITICALITY</th>
<th>COST</th>
<th>LOW</th>
<th>MODERATE</th>
<th>HIGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOW</td>
<td>LL</td>
<td>LM</td>
<td>LH</td>
<td></td>
</tr>
<tr>
<td>MODERATE</td>
<td>ML</td>
<td>MM</td>
<td>MH</td>
<td></td>
</tr>
<tr>
<td>HIGH</td>
<td>HL</td>
<td>HM</td>
<td>HH</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1.4 – Example of loss matrix to determine ordering policies*
Another way to determine the criticality of spare parts is discussed by Prakash et al. 1994, p.293-297 and Chen et al. 2009, p.226-228. They use an analytic hierarchy process (AHP) method to evaluate the criticality of spare parts. These other discussed classifications (in terms of cost and criticality) are important to decide which sets of spare parts to analyse and which not; it is obvious that a business will dedicate more efforts, time and money in the analysis of spare parts of high levels of cost and criticality rather than in the analysis of that with low levels.

3.1. Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) is a structured technique for dealing with complex decisions. Rather than prescribing a "correct" decision, the AHP helps the decision makers find the one that best suits their needs and their understanding of the problem.

Users of the AHP first decompose their decision problem into a hierarchy of more easily comprehended sub-problems, each of which can be analyzed independently. The elements of the hierarchy can relate to any aspect of the decision problem — tangible or intangible, carefully measured or roughly estimated, well- or poorly-understood — anything at all that applies to the decision at hand.

Once the hierarchy is built, the decision makers systematically evaluate its various elements by comparing them to one another two at a time. In making the comparisons, the decision makers can use concrete data about the elements, or they can use their judgments about the elements' relative meaning and importance. It is the essence of the AHP that human judgments, and not just the underlying information, can be used in performing the evaluations.

The AHP converts these evaluations to numerical values that can be processed and compared over the entire range of the problem. A numerical weight or priority is derived for each element of the hierarchy, allowing different and often incommensurable elements to be compared to one another in a rational and consistent way. This capability distinguishes the AHP from other decision making techniques.

In the final step of the process, numerical priorities are calculated for each of the decision alternatives. These numbers represent the alternatives' relative ability to achieve the decision goal, so they allow a straightforward consideration of the various courses of action.

The procedure for using the AHP can be summarized as:

1. Model the problem as a hierarchy containing the decision goal (criticality), the alternatives for reaching it (the spare parts), and the criteria for evaluating the alternatives (for example, purchase cost, cost of lack, required space).
2. Establish priorities among the criteria of the hierarchy (giving the weights).
3. Make a series of judgments based on pair-wise comparisons of the elements:
   for each criteria build a matrix with the alternatives both in lines and in columns,
   and numerical comparisons of dominance in the cells.
4. Determine the local weights for each couple criteria-alternative and calculate for
   each alternative the decision index by multiplying local weights with criteria
   weights and summing them.
5. Come to a final decision based on the results of this process.

4. Annexed costs
Different kinds of costs can be associated to spare parts management. The first cost is
the cost of lack: if there is a breakdown and no spare parts in the warehouse, there is a
cost associated to the loss of production which can be seen as missed pay-off.
Because of the complexity reached by the present productive systems, this costs can
be very significant. Sometimes, when the down times are high, some business are
induced to get on components that aren’t adapt; in this case there is the risk to damage
the productive system and to add other costs: reparation costs and further lack costs.
It is evident that connected to the storage of technical material such the spare parts
are, there is a significant financial cost which, in case of missed use of the item,
produces numerous negative effects. This financial cost includes the block of sums of
money for the purchase, the maintenance cost and eventually disposal cost in case of
missed utilize and turned up obsolescence (often due to the necessity to replace the
original productive system).
In conclusion, in the spare parts management for productive systems two contrasting
aspects have to be considered: the cost of lack and the cost of storage. The formulas
approved by the international literature to calculate this two kinds of costs are the
following:

\[
C_{\text{lack}} = P_{\text{lack}} \cdot \frac{T}{\text{MTTF}} \cdot C_h \cdot \text{MTTR} \quad (1.4)
\]

Where:
- \( P_{\text{lack}} \) is the probability of lack
- \( \text{MTTF} \) is the mean time to failure
- \( T \) is the interval time considered
- \( C_h \) is hourly cost of lack of production
- \( \text{MTTR} \) is the mean time to repair or replace
\[ C_{\text{storage}} = R \cdot t \cdot \bar{S} \]  

(1.5)

Where:
- \( R \) is the purchase cost of a spare part
- \( t \) is the financial storage rate
- \( S \) is the average storage of spare parts

The Figure 1.5 exemplifies the contrasting trend of the two costs in function of the level of supply of a spare part and the consequent trend of the total cost.

All the politics in the field of spare parts management have the same objective: to find the safe inventory level in order to minimize this total cost. In the present literature, subject and research on spare parts management mostly focus on the consideration of safe inventory level. Chen et al. (2009, p.225) say that if the actual required number of spare parts can be correctly predicted, there will be no problem of controlling inventory level and purchasing quantities.

Therefore, the following two chapter deal with the methods of forecasting this actual required number: chapter 2 presents an overview of recent literature in the field of spare parts demand forecasting methods, while chapter 3 presents methods based on neural networks.
CHAPTER 2
An overview of literature on spare parts demand forecasting methods

1. Introduction
In general there are three kinds of policies in the field of maintenance:

- Breakdown maintenance: replacement or repair is performed only at the time of failure. This may be the appropriate strategy in some cases, such as when the failure has no serious cost or safety consequences or it is low on the priority list.
- Preventive maintenance, where maintenance is performed on a scheduled basis with scheduled intervals often based on manufacturers’ recommendations and past experience of the equipment. This may involve replacement or repair, or both.
- Condition-based maintenance, where maintenance decisions are based on the current condition of the equipment, thus avoiding unnecessary maintenance and performing maintenance activities only when they are needed to avoid failure. CBM relies on condition monitoring techniques such as oil analysis, vibration analysis and other diagnostic techniques for making maintenance decisions.

Recent and non-recent scientific literature present a lot of maintenance policies, everyone based on one of the three presented kinds. They are different but they all have the same objective: to minimize the total cost. It is not scope of this work to explain these maintenance policies. The scope of this work is to give and discuss a series of demand forecasting methods; the spare parts demand forecasting or, to be more precise, the forecast of the number of breakdowns in a given period T, is the base of start of all the three kinds of maintenance policies, in particular breakdown and preventive maintenance.

Once the business know the actual number of required spare parts in a given period T, it is more sample to establish the safe inventory level.

2. Forecasting methods in the scientific literature
Because future demand plays a very important role in production planning and inventory management of spare parts, fairly accurate forecasts are needed. The manufacturing sector has been trying to manage the uncertainty of demand of spare parts...
parts for many years, which has brought about the development of many forecasting methods and techniques. Classical statistical methods, such as exponential smoothing and regression analysis, have been used by decision makers for several decades in forecasting spare parts demand. In addition to ‘uncertainty reduction methods’ like forecasting, ‘uncertainty management methods’ such as adding redundant spare parts have also been devised to cope with demand uncertainty in manufacturing planning and control systems (Bartezzaghi et al., 1999, p.501). Many of these uncertainty reduction or management methods may perform in a good way when CV is low, but in general perform poorly when demand for an item is lumpy or intermittent (Gutierrez et al., 2008, p.409). Lumpy demand has been observed in the automotive industry (Syntetos and Boylan, 2001,p.461-465; Syntetos and Boylan, 2005, p.310-313), in durable goods spare parts (Kalchschmidt et al., 2003, p.400-402), in aircraft maintenance service parts (Ghobbar and Friend, 2003, p.398), and in telecommunication systems, large compressors, and textile machines (Bartezzaghi et al., 1999, p.500), among others.

Croston (1972, p.290-293) was the first to note that, while single exponential smoothing has been frequently used for forecasting in inventory control systems, demand lumpiness generally leads to stock levels that are inappropriate. He noted a bias associated with placing the most weight on the most recent demand date, leading to demand estimates that tend to be highest just after a demand occurrence and lowest just before one. To address this bias, Croston proposed a new method of forecasting lumpy demand, using both the average size of nonzero demand occurrences and the average interval between such occurrences. Johnston and Boylan (1996, p.115-117) revisited Croston’s method, using simulation analysis to establish that the average inter-demand interval must be greater than 1.25 forecast revision periods in order for benefits of Croston’s method over exponential smoothing to be realized. On the other hand, Syntetos and Boylan (2001, p.460-461) reported an error in Croston’s mathematical derivation of expected demand and proposed a revision to approximately correct a resulting bias built into estimates of demand. Syntetos and Boylan (2005, p.304) quantified the bias associated with Croston’s method and introduced a new modification involving a factor of \((1-a/2)\) applied to Croston’s original estimator of mean demand, where \(a\) is the smoothing constant in use for updating the inter-demand intervals. This modification of Croston’s method, which has come to be known as the Syntetos–Boylan approximation (SBA), yields an approximately unbiased estimator. Syntetos and Boylan (2005, p.310-313) applied four forecasting methods - simple moving average over 13 periods, single exponential smoothing, Croston’s method, and SBA - on monthly lumpy demand histories (over a 2-year period) of 3000 stock keeping
units in the automotive industry. They undertook extended simulation experiments establishing the superiority of SBA over the three other methods, using relative geometric root-mean-square error as ordering criterion. Few years after the introduction of Croston’s method, Box and Jenkins (1976) introduce an iterative way to manage spare parts forecasts with ARMA, ARIMA, S-ARIMA methods that is nowadays largely used.

Ghobbar and Friend (2003, p.2105-2112) evaluated some of the above methods and other methods to forecast intermittent demand of aircraft’s spare parts. They compared and evaluated 13 methods, i.e. additive winter, multiplicative winter, seasonal regression model, component service life, weighted calculation of demand rates, weighted regression demand forecasters, Croston, single exponential smoothing, exponentially weighted moving average, trend adjusted exponential smoothing, weighted moving averages, double exponential smoothing, and adaptive-response-rate single exponential smoothing. Their results suggested that exponential smoothing and Croston’s methods outperformed other forecasting methods for intermittent demand.

Also methods based on Poisson distribution have been studied in the field of spare parts (Manzini et al., 2007, p.205-212). Hill et al. (1996, p.1083-1084) pointed out that traditional statistical time-series methods can misjudge the functional form relating the independent and dependent variables. These misjudged relationships are inflexible to modification during the model building process. These traditional methods can also fail to make necessary data transformations. For this, in the very last years an innovative method based on human intelligence has captured the attention of the experts in this field, mainly for the experimentally results that it has reached: artificial neural network. Gutierrez et al.(2008, p.409-410) say that traditional time-series methods may not sometimes capture the nonlinear pattern in data. Artificial neural network (ANN) or, simply, neural network (NN) modelling is a logical choice to overcome these limitations. NN models can provide good approximations to just about any functional relationships. Successful applications of NNs have been well documented as early as in the 1980s.

Another forecasting method (grey prediction model) has been introduced in the very last years and has been valued well-performing in particular for forecasts of low term (Tzeng et al., 2004, p.5). In brief, a lot of forecasting methods have been elaborated and studied in the filed of spare parts demand. The following two tables give a panoramic of the methods that have been more discussed than others or have been valued as the most successful by the scientific literature: the first table gives a brief description, with also limits and advantages, of every method that is after better explained, while the second table shows the most recent or most studied articles that deal with every method.
<table>
<thead>
<tr>
<th>METHODS</th>
<th>ABBR.</th>
<th>INPUTS</th>
<th>DESCRIPTION</th>
<th>MATHEMATIC MODEL</th>
<th>INNOVATIVE FEATURES</th>
<th>LIMITS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINGLE EXPONENTIAL SMOOTHING</td>
<td>SES</td>
<td>- historical data</td>
<td>It adopts a smoothing constant $\alpha$ of the real demands</td>
<td>- Exponential</td>
<td>- adapt for low-period forecasts</td>
<td>- Deterministic model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- smoothing constant</td>
<td></td>
<td>smoothing</td>
<td>- easy to compute</td>
<td>- few fields of applicability</td>
</tr>
<tr>
<td>CROSTON’S METHOD</td>
<td>Croston</td>
<td>- historical data</td>
<td>Evolution of SES which also looks on intervals of zero demand</td>
<td>- Exponential</td>
<td>- adapt to demand with a lot of zero values</td>
<td>- Deterministic model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- interval between present and last non-zero demand</td>
<td></td>
<td>smoothing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- smoothing constant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYNTETOS – BOYLAN APPROXIMATION</td>
<td>SBA</td>
<td>- historical data</td>
<td>Evolution of Croston in order to decrease the error of the expected estimate of demand per time period</td>
<td>- Exponential</td>
<td>- decrease of the theoretical error of Croston’s method</td>
<td>- Deterministic model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- interval between present and last-non zero demand</td>
<td></td>
<td>smoothing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- smoothing constant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOVING AVERAGE</td>
<td>MA</td>
<td>- historical data</td>
<td>Mean of the past $n$ demands</td>
<td>- Arithmetic mean</td>
<td>- adapt for the constant demands</td>
<td>- Deterministic model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- number of data to considerate</td>
<td></td>
<td></td>
<td>- easy to compute</td>
<td>- Few fields of applicability</td>
</tr>
<tr>
<td>WEIGHTED MOVING AVERAGE</td>
<td>WMA</td>
<td>- historical data</td>
<td>Mean of past $n$ demands with decreasing weights</td>
<td>- Arithmetic mean</td>
<td>- more weight applied to last demands</td>
<td>- Deterministic model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- number of data to considerate</td>
<td></td>
<td></td>
<td>- easy to compute</td>
<td>- applicable only with low level of lumpiness</td>
</tr>
<tr>
<td>METHODS</td>
<td>ABBR.</td>
<td>INPUTS</td>
<td>DESCRIPTION</td>
<td>MATHEMATIC MODEL</td>
<td>INNOVATIVE FEATURES</td>
<td>LIMITS</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------</td>
<td>------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ADDITIVE WINTER</td>
<td>AW</td>
<td>- historical data</td>
<td>Evolution of SES with the introduction of additive terms on the components (trend, casual component, ...)</td>
<td>- exponential smoothing</td>
<td>- it considers the effects of seasonality (in a additive way)</td>
<td>- in few fields spare parts deal with seasonality - deterministic model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- smoothing constant $\alpha$</td>
<td></td>
<td>- sum of components</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- trend constant $\beta$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- periodicity constant $\gamma$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- width of periodicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MULTIPLICATIVE WINTER</td>
<td>MW</td>
<td>- historical data</td>
<td>Evolution of SES with the introduction of multiplicative terms on the components (trend, casual component, ...)</td>
<td>- exponential smoothing</td>
<td>- it considers the effects of seasonality (in a multiplicative way)</td>
<td>- in few fields spare parts deal with seasonality - deterministic model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- smoothing constant $\alpha$</td>
<td></td>
<td>- product of components</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- trend constant $\beta$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- periodicity constant $\gamma$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- width of periodicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOOTSTRAP METHOD</td>
<td>Boot</td>
<td>- historical data</td>
<td>Modern approach to statistical inference, falling within a broader class of re-sampling methods</td>
<td>- probabilistic model (re-sampling)</td>
<td>- it a values the demand in a probabilistic way - adapt in case of limited historical data</td>
<td>- sometimes it can lead to extremely biased forecast</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- number of re-sampling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- width of a sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POISSON METHOD</td>
<td>Poisson PM</td>
<td>- historical data</td>
<td>Application of the binomial formula to forecast</td>
<td>- probabilistic model (binomial distribution)</td>
<td>- it a values the demand in a probabilistic way - adapt in case of rare demand</td>
<td>- it doesn’t give a punctual value - over-estimated forecasts in erratic or lumpy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- interval time $T$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- punctual value of the $x$ demand to forecast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>METHODS</td>
<td>ABBR.</td>
<td>INPUTS</td>
<td>DESCRIPTION</td>
<td>MATHEMATIC MODEL</td>
<td>INNOVATIVE FEATURES</td>
<td>LIMITS</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------</td>
<td>-------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>-----------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>BINOMIAL METHOD</td>
<td>BM</td>
<td>- historical data</td>
<td>It values the forecast demand as sum of two terms, associated at the probability of happening</td>
<td>- probabilistic model (binomial distribution)</td>
<td>- it a values the demand in a probabilistic way</td>
<td>- it give not the exact forecast, but a number of spare parts to guarantee a level of service</td>
</tr>
<tr>
<td>GTREY PREDICTION MODEL</td>
<td>GM</td>
<td>- historical data</td>
<td>With probabilistic basis, this algorithm forecasts through the use of cumulative demand and least square method to minimize the error</td>
<td>- accumulative generating operation</td>
<td>- Ideal when there are few historical data - it performs good in low-period forecasts</td>
<td>- not well performing in the medium and long term</td>
</tr>
<tr>
<td>ARMA ARIMA S-ARIMA (BOX- JENKINS METHODS)</td>
<td>BJ</td>
<td>- historical data</td>
<td>They combine autoregressive and moving average models in an iterative way until the best forecasts are produced</td>
<td>- autoregression - weighted average of residuals</td>
<td>- possibility to consider non-stationarity and seasonality - iterative way until best performances</td>
<td>- they requires a lot of historical data to give good results</td>
</tr>
<tr>
<td>NEURAL NETWORK</td>
<td>NN or ANN</td>
<td>- historical data</td>
<td>Based on human intelligence, it learns from a training set the connection between inputs and output (the forecast)</td>
<td>Not mathematical model</td>
<td>- it learns a automatically the connections between output and inputs</td>
<td>- it requires a lot of historical data to give good results - not easy to compute</td>
</tr>
<tr>
<td>AUTHORS</td>
<td>YEAR</td>
<td>SES</td>
<td>CR</td>
<td>SBA</td>
<td>MA</td>
<td>WMA</td>
</tr>
<tr>
<td>------------------------</td>
<td>------</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>Amin-Naseri, Tabar</td>
<td>2008</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archibald , Koehler</td>
<td>2003</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bookbinder , Lordahl</td>
<td>1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bu Hamra et al.</td>
<td>2003</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen F.L., Chen Y.C.</td>
<td>2009</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen F.L., Chen Y.C.</td>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Croston</td>
<td>1972</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghobbar A.A., Friend</td>
<td>2003</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Gutierrez et al.</td>
<td>2008</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill et al.</td>
<td>1996</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ho, Xie</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ho et al.</td>
<td>2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hua, Zhang</td>
<td>2006</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu, Li</td>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnston , Boylan</td>
<td>1996</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koehler et al.</td>
<td>2001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lawton</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mckenzie</td>
<td>1986</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramjee , Crato</td>
<td>2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rao</td>
<td>1973</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUTHORS</td>
<td>YEAR</td>
<td>SES</td>
<td>CR</td>
<td>SBA</td>
<td>MA</td>
<td>WMA</td>
</tr>
<tr>
<td>------------------------</td>
<td>------</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>Sani, Kingsman</td>
<td>1997</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schulz</td>
<td>1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheu et al.</td>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snyder</td>
<td>2002</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntetos, Boylan</td>
<td>2001</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntetos, Boylan</td>
<td>2005</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntetos et al.</td>
<td>2005</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teunter, Sani</td>
<td>2009</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teunter et al.</td>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang, Rao</td>
<td>1992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tseng et al.</td>
<td>2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tzeng et al.</td>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willemain et al.</td>
<td>2004</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willemain et al.</td>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xu et al.</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yar, Chatfield</td>
<td>1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A - Comparative evaluations based on experimental data  
B - Individuation of contexts of application  
C - Theoretical explanation  
D - Proposition of innovative elements  
E - Individuation of errors in the method  
F - Experimental application
3. Explanation of the forecasting methods

3.1. Single exponential smoothing
This method is based on time series analysis, particularly adapt for low period forecast. In substance, the forecast of spare parts demand is obtained by applying a series of weights, decreasing in an exponential way, at the historical data. The forecast formula is this:
\[ F_{t+1} = \alpha X_t + (1 - \alpha) F_t \]  
(3.1)
Where:
\( X_t \) is the actual value of the demand at the instant \( t \);
\( F_{t+1} \) is the forecast for instant \( t+1 \);
\( \alpha \) is the smoothing parameter.
Smoothing parameter \( \alpha \) can have different values, generally between 0.1 and 0.4 on the basis of demand features (with unstable demand high values for the parameter are used).

3.2. Croston's method
Croston proposed a method (abbreviated as CR) that takes account of both demand size and inter-arrival time between demands. The method is now widely used in industry and it is incorporated in various best selling forecasting software packages. The CR method has been assessed by several authors since 1972.
Rao (1973, p.639-640) made corrections to several expressions in Croston’s paper without affecting the final conclusions or the forecasting procedure. Schultz (1987, p.454-457) presented a forecasting procedure, which is basically the CR method and suggested a base-stock inventory policy with replenishment delays. He proposed the use of two smoothing parameters (one for demand size, the other for demand intervals), whereas in the original paper by Croston (1972, p.291-295) a common smoothing parameter was assumed. Willemain et al. (1994, p.530-535) compared the CR method with exponential smoothing and concluded that the CR method is robustly superior to exponential smoothing, although results with real data in some cases show a more modest benefit. Johnston and Boylan (1996, p.115-120) obtained similar results, but further showed that the CR method is always better than exponential smoothing when the average inter-arrival time between demands is greater than 1.25 review intervals. Sani and Kingsman (1997, p.705-710) compared various forecasting and inventory control methods on some long series of low demand real data from a typical spare parts depot in the UK. They concluded, based on cost and service level,
that the best forecasting method is moving average followed by the CR method. An important contribution is that by Syntetos and Boylan (2001, p.458-465). They showed that the CR method leads to a biased estimate of demand per unit time. They also propose a modified method (SBA) and demonstrate the improvement in a simulation experiment. Snyder (2002, p.686-692) critically assessed the CR method with a view to overcome certain implementation difficulties on the data sets used. Snyder made corrections to the underlying theory and proposed modifications.

An important study was conducted by Teunter et al. (2010, p.179-182): after having showed in a previous article (using a large data set from the UK’s Royal Air Force) that CR and SBA all outperform moving average and exponential smoothing, they compared in a numerical example Croston method and all his different variations proposed in the years.

Croston’s original method (CR) forecasts separately the time between consecutive transactions \(P_t\) and the magnitude of the individual transactions \(Z_t\). At the review period \(t\), if no demand occurs in a review period then the estimates of the demand size and inter-arrival time at the end of time \(t\), \(Z_t\) and \(P_t\), respectively, remain unchanged. If a demand occurs so that \(X_t > 0\), then the estimates are updated by:

\[
Z_t = \alpha \cdot X_t + (1 - \alpha) \cdot Z_{t-1} \quad (3.2)
\]

\[
P_t = \alpha \cdot G_t + (1 - \alpha) \cdot P_{t-1} \quad (3.3)
\]

Where:

- \(X_t\) actual value of the demand at the instant \(t\);
- \(G_t\) actual value of the time between consecutive transactions at the instant \(t\);
- \(\alpha\) smoothing constant between zero and one.

Hence, the forecast of demand per period at time \(t\) is given as:

\[
F_{t+1} = \frac{Z_t}{P_t} \quad (3.4)
\]

3.3. Syntetos – Boylan Approximation

An error in Croston’s mathematical derivation of expected demand size was reported by Syntetos and Boylan (2001, p.459-461), who proposed a revision to approximately correct Croston’s demand estimates: the SBA or SB method.

In an attempt to confirm the good performance of their SB method, Syntetos and Boylan (2005, p.309-313) carried out a comparison of forecasting methods including theirs and the original CR method. A simulation exercise was carried out on 3000 products from the automotive industry with “fast intermittent” demand. It was shown that the modification is the most accurate estimator. In another study, Syntetos et al.
(2005, p.497-502) analyzed a wider range of intermittent demand patterns and made a categorisation to guide the selection of forecasting methods. They indicated that there are demand categories that are better used with the CR method and there are others that go well with the SBA method.

There are several variation applied at Croston’s method after his introduction in 1972, and SBA is considered one the most performing by several authors.

Syntetos and Boylan (2001, p.459-460) pointed out that Croston’s original method is biased. They showed that in CR the expected value is not $\mu/p$, but:

$$E(F_t) = \frac{\mu}{p} \left( 1 + \frac{\alpha}{2 - \alpha} \cdot \frac{p-1}{p} \right) \quad (3.5)$$

Where:

$\mu$ is the mean of historical demand;

$p$ is the mean of historical inter-demand intervals $P_t$.

And, in particular, for $\alpha = 1$:

$$E(F_t) = \mu \cdot \left[ -\frac{1}{p-1} \cdot \ln \left( \frac{1}{p} \right) \right] \quad (3.6)$$

Based on 3.5 and ignoring the term $(p-1)/p$, Syntetos and Boylan proposed a new estimator given as:

$$F_{t+1} = \left( 1 - \frac{\alpha}{2} \right) \frac{Z_t}{P_t} \quad (3.7)$$

One can expect this new estimator to perform better as $(p-1)/p$ gets closer to one, i.e., as the probability $1/p$ of positive demand in a period gets smaller. The effect is that Croston’s original method has a smaller (positive) bias if $1/p$ is large (few demands are zero), and the Syntetos - Boylan modification has a smaller bias if $1/p$ is small (many demands are zero).

3.4.Moving Average
The moving average (MA) method is the mean of the previous $n$ data sets. The formula for the moving average is:

$$F_t = MA(n) = \frac{X_{t-1} + X_{t-2} + \ldots + X_{t-n}}{n} \quad (3.8)$$

As it transpires from the formula, this method is really simple and easy to compute, but it is applicable only in case of slow moving demand. In the other cases the demand gravitates with difficulty around the average of last $n$ periods.
3.5. Weighted moving average

A weighted average is any average that has multiplying factors to give different weights to different data points. Mathematically, the moving average is the convolution of the data points with a moving average function; in technical analysis, a WMA has the specific meaning of weights that decrease arithmetically. In an \( n \)-period WMA the latest period has weight \( n \), the second latest \( n-1 \), etc, down to zero.

\[
F_{t+1} = \frac{n \cdot p_t + (n-1) \cdot p_{t-1} + \ldots + 2 \cdot p_{(t-n)+2} + p_{(t-n)+1}}{n + (n-1) + \ldots + 2 + 1}
\]  

(3.9)

The graph below shows how the weights decrease, from highest weight for the most recent data points, down to zero.

![Fig 3.1 - WMA weights \( n = 15 \)](image)

This is an example of WMA; in general WMA is any average with different weights applied to past values of demand.

3.6. Holt–Winters methods

Additive and multiplicative winter are two methods proposed by Winters and Holt in order to considerate hypothetical seasonal effects. A first way to considerate these seasonal effects is the introduction of a drift \( D \) which modifies the levelled values according to variables which depend upon time. Drift \( d \) is a function which represents the trend. For example, a model which considerate trend effect is this:

\[
F_{t+k} = L_t + D_t \cdot k
\]  

(3.10)

with the following relations:

\[
L_t = \alpha \cdot y_t + (1-\alpha) \cdot (L_{t-1} + D_{t-1})
\]  

(3.11)

\[
D_t = \beta \cdot (L_t - L_{t-1}) + (1-\beta) \cdot D_{t-1}
\]  

(3.12)

The first can be seen as a weighted average of the observed value \( y_t \) and the forecast calculated at the previous period; the second as a weighted average of the difference
between forecasts calculated at the period t and t-1 and the drift calculated at the period t-1 (to attribute a weight \( \beta \) equal to 1 to this last one is equivalent to assume a linear trend, that is a constancy in the drift).

The AW and MW are an extension of this first example in order to also consider the seasonality in strict meaning. The Additive Winter starts from the following relations:

\[
L_t = \alpha \cdot (y_t - S_{t-p}) + (1 - \alpha) \cdot (L_{t-1} + D_{t-1}) \tag{3.13}
\]

\[
D_t = \beta \cdot (L_t - L_{t-1}) + (1 - \beta) \cdot D_{t-1} \tag{3.14}
\]

\[
S_t = \gamma \cdot (y_t - L_{t-1}) - (1 - \gamma) \cdot S_{t-p} \tag{3.15}
\]

where \( s_t \) is a factor of seasonality and \( p \) his periodicity (4 for quarterly data, 12 for monthly data, and so on). The demand forecast for the period \( t \) is:

\[
F_{t+k} = L_t + D_{t+k} + S_{t+k-p} \tag{3.16}
\]

In parallel, Multiplicative Winter has the following relations:

\[
L_t = \alpha \cdot \frac{y_t}{S_{t-p}} + (1 - \alpha) \cdot (L_{t-1} + D_{t-1}) \tag{3.17}
\]

\[
D_t = \beta \cdot (L_t - L_{t-1}) + (1 - \beta) \cdot D_{t-1} \tag{3.18}
\]

\[
S_t = \gamma \cdot \frac{y_t}{L_t} + (1 - \gamma) \cdot S_{t-p} \tag{3.19}
\]

and the forecast demand for the period \( t \) is:

\[
F_{t+k} = (L_t + D_{t+k}) \cdot S_{t+k-p} \tag{3.20}
\]

These models are very flexible, because they can also consider non-polynomial trends and not-constant seasonality. With regard to the choice of the weights \( \alpha \), \( \beta \) and \( \gamma \), values the minimize the square of the gaps can be taken or, in alternative, they can be chosen in line with the scope of the analysis.

### 3.7. Bootstrap method

Hua et al. (2006, p.1037) say that when historical data are limited, the bootstrap method is a useful tool to estimate the demand of spare parts. Bookbinder and Lordahl (1989, p 303) found the bootstrap superior to the normal approximation for estimating high percentiles of spare parts demand for independent data. Wang and Rao (1992, p 333-336) also found the bootstrap effective to deal with smooth demand. All these papers do not consider the special problems of managing intermittent demand. Willemain et al. (2004, p.377-381) provided an approach of forecasting intermittent demand for service parts inventories. They developed a bootstrap-based approach to forecast the distribution of the sum of intermittent demands over a fixed lead time.
Bootstrapping is a modern, computer-intensive, general purpose approach to statistical inference, falling within a broader class of re-sampling methods. Bootstrapping is the practice of estimating properties of an estimator (such as its variance) by measuring those properties when sampling from an approximating distribution. One standard choice for an approximating distribution is the empirical distribution of the observed data. In the case where a set of observations can be assumed to be from an independent and identically distributed population, this can be implemented by constructing a number of re-samples of the observed dataset (and of equal size to the observed dataset), each of which is obtained by random sampling with replacement from the original dataset.

The bootstrap procedure can be illustrated with the following steps:

1- take an observed sample (in our case a sample of historical spare parts demand) of number equal to \( n \), called \( X = (x_1, x_2, \ldots, x_n) \);
2- from \( X \), resample \( m \) other samples of number equal to \( n \) obtaining \( X_1, X_2, \ldots, X_m \) (in every bootstrap extraction, the data of the observed sample can be extracted more than one time and every data has the probability \( 1/n \) to be extracted);
3- given \( T \) the estimator of \( \theta \), parameter of study (in our case it may be the average demand), calculate \( T \) for every bootstrap sample. In this way we have \( m \) estimates of \( \theta \);
4- from these estimates calculate the desired value: in our case the mean of \( T_1, \ldots, T_m \) can be the demand forecast.

This method can be applied not only to find the average demand (that can be the demand forecast) but also the intervals between non-zero-demand or other desired values.

### 3.8 Poisson method

Poisson method is typically used for the forecast of the probability of happening of a rare event (Manzini et al., 2007, p.205). It derives directly from the binomial distribution. This method doesn’t allow the direct calculation of the variable to forecast, but it consents an estimate of the probability that it assumes a determined value.

The point of start of this model is the valuation of the average value of the variable to forecast. In case of spare parts, given the average consumption in an interval time \( T \) equal to \( d \), the probability to have a demand equal to \( x \) (i.e. \( x \) requires of components) in the interval time \( T \) is:
\[ P_{d,T,x} = \frac{(d \cdot T)^x \cdot e^{-(d \cdot T)}}{x!} \] (3.21)

In consequence, the cumulative probability (a measure that not more than \( x \) components are required) can be expressed as:

\[ P_{\text{CUM}d,T,x} = \sum_{k=0}^{x} \frac{(d \cdot T)^k \cdot e^{-(d \cdot T)}}{k!} \] (3.22)

### 3.9. Binomial method

This method was introduced as evolution of the application of Poisson formula. Effectively, with Poisson model there often are inaccurate forecasts (by nature overestimated), above all in case of erratic and lumpy demand (Manzini et al., 2007, p.209). Binomial method values through a model composed by two additive terms the demand of a spare part, having as point of departure the average consumption of the item. The method can also consider the eventual simultaneous use of a single type of spare parts in several applications, through the parameter \( n \).

The forecast formula is the following:

\[ N = x_1 + x_2 \] (3.23)

with: \[ x_1 = \frac{T}{(1/d)} \cdot n \] (3.24)

where:

- \( N \) is the forecast demand
- \( d \) is the historical average consumption
- \( T \) is the interval time considered for the estimation of the requirements

The term \( x_2 \) is defined in connection with the accepted probability that exactly \( x_2 \) breakdowns happen in the interval time \( T \), defining \( T_{\text{residual}} \) the time “not covered” by mean term \( x_1 \).

\( T_{\text{residual}} \) is defined as follow:

\[ T_{\text{residual}} = T - \frac{T}{(1/d)} \cdot (1/d) \] (3.25)

At this point cumulative probability of consumption \( p \) is introduced in the period \( T_{\text{residual}} \) assuming an exponential function:

\[ F(T_{\text{residual}}) = 1 - e^{-(1/Td) \cdot T_{\text{residual}}} = p \] (3.26)

A level of service – LS that has to be assured in the \( T_{\text{residual}} \) period (in other term LS represents the probability with which to cover the eventual spare part demand in the fixed interval time) is fixed.
Taking advantage of the properties of the binomial formula, through an iterative procedure, the value of $x_2$ spare parts that allow to achieve the desired LS can be determined. In other terms, this means to find the value of $x_2$ for which $P(x_2)$ is major than LS:

$$P(x_2) = \sum_{i=0}^{x_2} \binom{n}{i} (1-p)^{n-i} \cdot p^i \geq \text{LS} \quad (3.27)$$

The term $x_1$ represents an average value which produces reliable forecasts in case of high average demand. Because of, on the contrary, the consumption of spare parts are often of low levels, it is opportune to not disregard the decimal part: this is the “spirit” that generates the term $x_2$.

The two last considered methods (Poisson and binomial method) don’t seem to be based on historical data. In reality, also in these cases, historical data are very important for the calculation of the variable $d$.

3.10 Grey prediction model

Grey theory, originally developed in the 80-years, focuses on model uncertainty and information insufficiency in analyzing and understanding systems via research on conditional analysis, prediction and decision-making.

Grey forecasting differs from other statistical regression models. With a basis in probability theory, conventional regression requires amount of data for establishing forecast model. Grey forecasting is based on the grey generating function (GM(1,1) model is the most frequently used grey prediction method), which uses the variation within the system to find the relations between sequential data and establish then the prediction model.

The procedure of GM (1, 1) grey prediction model can be summarized as follows.

Step 1. Establish the initial sequence from observed data

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2),..., x^{(0)}(n))$$

where $x^{(0)}(i)$ represents the base line (state = 0) data with respect to time $i$.

Step 2. Generate the first-order accumulated generating operation (AGO) sequence

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2),..., x^{(1)}(n))$$

where $x^{(1)}(k)$ is derived as following formula:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(1)}(i) \quad (3.28)$$

Step 3. Compute the mean value of the first-order AGO sequence:
\[ Z^{(1)}(k) = 0.5 \cdot x^{(1)}(k) + 0.5 \cdot x^{(1)}(k - 1) \]  \hspace{1cm} (3.29)

Step 4. Define the first-order differential equation of sequence \( x(1) \) as:

\[ \frac{dx^{(1)}(k)}{dk} + ax^{(1)}(k) = b \]  \hspace{1cm} (3.30)

where \( a \) and \( b \) express the estimated parameters of grey forecasting model.

Step 5. Utilizing the least squares estimation, we can derive the estimated first-order AGO sequence \( x^{(1)}(k+1) \) and the estimated inversed AGO sequence \( x^{(0)}(k+1) \) (the forecast) as follows:

\[ x^{(1)}(k+1) = \left[ x^{(0)}(k) - \frac{b}{a} \right] \cdot e^{-ak} + \frac{b}{a} \]  \hspace{1cm} (3.31)

\[ x^{(0)}(k+1) = x^{(1)}(k + 1) - x^{(1)}(k) \]  \hspace{1cm} (3.32)

where parameter \( a \) and \( b \) can be conducted by following equations:

\[
\begin{bmatrix}
a \\
b
\end{bmatrix} = (B^T \cdot B)^{-1} \cdot B^T \cdot y
\]  \hspace{1cm} (3.33)

\[
B = \begin{bmatrix}
-0.5 \cdot (x^{(1)}(1) + x^{(1)}(2)) & 1 \\
-0.5 \cdot (x^{(1)}(2) + x^{(1)}(3)) & 1 \\
& \ldots \\
-0.5 \cdot (x^{(1)}(n - 1) + x^{(1)}(n)) & 1
\end{bmatrix}
\]  \hspace{1cm} (3.34)

\[ y = \begin{bmatrix}
x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n)
\end{bmatrix}^T
\]  \hspace{1cm} (3.35)

3.11..ARMA(p,q) ARIMA(p,d,q) S-ARIMA(p,d,q)(P,D,Q)s

This is a group of methods which consist of two parts: an autoregressive (AR) part and a moving average (MA) part.

An autoregressive model of order \( p \) has the form:

\[ F_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t \]  \hspace{1cm} (3.36)

where:

- \( u_i \) is the actual value in the period \( i \);
- \( \rho_i \) is a coefficient;
- \( \varepsilon_i \) is a residual term that represents random events not explained by model.

A moving average forecasting model, in this case, uses lagged values of the forecast error \( \varepsilon \) to improve the current forecast. A first-order moving average term uses the most recent forecast error, a second-order term uses the forecast error from the two most recent periods, and so on. An MA(q) has the form:

\[ F_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \]  \hspace{1cm} (3.37)

where:
- $\varepsilon_i$ is the residual of the period $i$;
- $\theta_i$ is a coefficient;

### 3.11.1. ARMA($p,q$)

This method is used when the time series is stationary (a stationary time series is one whose average is not changing over time).

The forecasting is formula is:

$$F_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \quad (3.38)$$

AR and MA are combined: $p$ is the degree of AR, and $q$ is the degree of MA.

Degrees $p$ and $q$ are chosen by analyzing the global and partial autocorrelation. The first measures, varying $k$, the relation between $u_t$ and $u_{t-k}$, also considering the variables $u_{t-1}, \ldots, u_{t-k+1}$. The second measures the relation between $u_t$ and $u_{t-k}$, without considering other variables. Global and partial autocorrelation are analyzed by the correlogram and the degrees $p$ and $q$ that have to be used are tied to the distribution shown by the correlogram; some examples are in Hanke and Reitsch, 1992, p.383-385.

### 3.11.2. ARIMA($p,d,q$)

An autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. It is applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to remove the non-stationarity. When the is removed the process is the same of ARMA.

The model is generally referred to as an ARIMA($p,d,q$) model where $p$, $d$, and $q$ are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. When one of the terms is zero, it's usual to drop AR, I or MA. For example, an I(1) model is ARIMA(0,1,0), and a MA(1) model is ARIMA(0,0,1).

### 3.11.3. S-ARIMA($p,d,q$)/($P,D,Q$)$s$

This method is used in case of seasonality of order $s$. The procedure is the same of ARIMA but in this case there are three other degrees: $P$, $D$ and $Q$; they have the same meaning of $p$, $d$, $q$ but only applied to the seasonal data in the periods $t$, $t-n$, $t-2n$, ..., where $n$ is the number of periods in the year divided by $s$. 


3.11.4. BOX-JENKINS METHODOLOGY

This procedure, gives a way to decide how to use these three forecasting models. This technique does not assume any particular pattern in the historical data of the series to be forecast. It uses an iterative approach of identifying a possible useful model from a general class of models. The chosen model is then checked against the historical data to see whether it accurately describes the series. The model fits well if the residuals between the forecasting model and the historical data points are small, randomly distributed, and independent. If the specified model is not satisfactory, the process is repeated by using another model designed to improve on the original one. This process is repeated until a satisfactory model is found. Figure 3. illustrates the approach.

![Box-Jenkins Procedure Diagram]

3.12. Neural networks

The application of neural networks in the field of spare parts are the centre of almost all scientific studies of the very last years. Artificial neural networks (ANN) are computing models for information processing and pattern identification. They grow out of research interest in modeling biological neural systems, especially human brains. An ANN is a network of many simple computing units called neurons or cells, which are highly interconnected and organized in layers. Each neuron performs the simple task of
information processing by converting received inputs into processed outputs. Through the linking arcs among these neurons, knowledge can be generated and stored regarding the strength of the relationship between different nodes. Although the ANN models used in all applications are much simpler than actual neural systems, they are able to perform a variety of tasks and achieve remarkable results. A detailed explanation of the theory of NN and their application in the field of spare parts demand forecasting are the objectives of chapter 3.

4. Benchmarks

Benchmarking, in this field, is the process of comparing different forecasting methods in order to determinate which has more confirmations in the reality. Benchmarks are the parameters, the references with which two or more forecasting methods are evaluated, in connection with the actual demands that occurred. In the scientific literature several types of benchmarks have been used; in the following paragraphs the most used will be explained. There are two kinds of parameters: absolute accuracy measures (4.1.– 4.2.– 4.3.– 4.4.) and accuracy measures relative to other methods (4.5.– 4.6.).

4.1. MAPE

Mean absolute percentage error (MAPE) expresses accuracy as a percentage, and is defined by the formula:

\[
MAPE = \frac{1}{n} \sum_{t=1}^{N} \left| \frac{A_t - F_t}{A_t} \right| \quad (4.1)
\]

where \(A_t\) is the actual value and \(F_t\) is the forecast value.

The difference between \(A_t\) and \(F_t\) is divided by the actual value \(A_t\) again. The absolute value of this calculation is summed for every fitted or forecast point in time and divided again by the number of fitted points \(n\). This makes it a percentage error so one can compare the error of fitted time series that differ in level.

Although the concept of MAPE sounds very simple and convincing, it has two major drawbacks in practical application:

- If there are zero values (which sometimes happens in spare parts demand series) there will be a division by zero.
- When having a perfect fit, MAPE is zero. But in regard to its upper level the MAPE has no restriction. When calculating the average MAPE for a number of time series there might be a problem: a few number of series that have a very high MAPE might distort a comparison between the average MAPE of time series.
series fitted with one method compared to the average MAPE when using another method. In order to avoid this problem other measures have been defined, for example the S-MAPE (symmetrical MAPE) or a relative measure of accuracy.

### 4.2. S-MAPE

Symmetric mean absolute percentage error (S-MAPE) is an accuracy measure based on percentage (or relative) errors. It is usually defined as follows:

\[
S-\text{MAPE} = \frac{1}{n} \sum_{t=1}^{N} \left| \frac{A_t - F_t}{(A_t + F_t)/2} \right|
\]

where \( A_t \) is the actual value and \( F_t \) is the forecast value.

The absolute difference between \( A_t \) and \( F_t \) is divided by half the sum of the actual value \( A_t \) and the forecast value \( F_t \). The value of this calculation is summed for every fitted point \( t \) and divided again by the number of fitted points \( n \).

Contrary to the mean absolute percentage error, SMAPE has a lower bound and an upper bound. Indeed, the formula above provides a result between 0% and 200%. However a percentage error between 0% and 100% is much easier to interpret. That is the reason why the formula below is often used in practice (i.e. no factor 0.5 in denominator):

\[
S-\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{A_t - F_t}{A_t + F_t} \right|
\]

However, one problem with S-MAPE is that it is not as symmetric as it sounds since over- and under-forecasts are not treated equally. Let’s consider the following example by applying the second S-MAPE formula:

- Over-forecasting: \( A_t = 100 \) and \( F_t = 110 \) give S-MAPE = 4.76%
- Under-forecasting: \( A_t = 100 \) and \( F_t = 90 \) give S-MAPE = 5.26%.

### 4.3. A-MAPE

Several variations of MAPE have been suggested in the scientific literature (an important work has been done by Hover, 2006, p.32-35), among which adjusted mean absolute percentage error (A-MAPE) is one of the most used in comparing spare parts demand forecasting methods. The formula is this:
The root mean square deviation (RMSD) or root mean square error (RMSE) is a frequently-used measure of the differences between values predicted and the values actually observed.

The formula is:

\[ \text{RMSD} = \frac{1}{N} \sum_{t=1}^{N} \sqrt{(F_t - A_t)^2} \]  

(4.5)

A study conducted by Armstrong and Collopy, 1992, p.69-80, evaluated measures for making comparisons of errors across 90 annual and 101 quarterly time-series data. The study concluded that MAPE should not be the choice if large errors are expected because MAPE is biased in favour of low forecasts. The study also concluded that root mean square error (RMSE) is not reliable, even though most practitioners prefer RMSE to all other error measures since it describes the magnitude of the errors in terms useful to decision makers (Carbone and Armstrong, 1982, p.215-217). The study recommended the adjusted mean absolute percentage error (A-MAPE) statistic for selecting the most accurate methods when many time-series data are available. However, computing A-MAPE for intermittent demand is difficult because of zero demand over many time periods.

4.5.RGRMSE

Syntetos and Boylan (2005, p.305-309) used two accuracy measures relative to other methods. The first measure, relative geometric root-mean-square error (RGRMSE), is given by:

\[ \text{RGRMSE} = \left( \dfrac{\prod_{t=1}^{N} (A_{a,t} - F_{a,t})^2}{\prod_{t=1}^{N} (A_{b,t} - F_{b,t})^2} \right)^{1/2n} \]  

(4.6)

where the symbols \( A_{k,t} \) and \( F_{k,t} \) denote actual demand and forecast demand, respectively, under forecasting method \( k \) at the end of time period \( t \). If RGRMSE is lower than 1, method \( a \) performs better than method \( b \). Fildes (1992, p.93-94) argued that RGRMSE has a desirable statistical property. According to him the error in a
particular time period consists of two parts: one due to the method and the other due to the time period only. RGRMSE expressed in a relative way is independent of the error due to the time period, thereby focusing only on the relative merits of the methods.

4.6. PB
The second error measure, the percentage best (PB), is the percentage of time periods one method performs better than the other methods under consideration. PB is particularly meaningful because all series and all data periods in each series generate results (Syntetos and Boylan, 2005, p.308). The mathematical expression for PB for method m is:

\[ PB_m = \frac{100}{N} \sum_{t=1}^{N} B_{m,t} \]  

(4.7)

where for time period t, \( B_{m,t} = 1 \) if \( |A_{m,t} - F_{m,t}| \) is the minimum of \( |A_{k,t} - F_{k,t}| \) for all methods k under consideration, and \( B_{m,t} = 0 \) otherwise.

In the evaluation of k methods, the method that has the greatest PB is the method which performs better.
CHAPTER 3
Neural networks in spare parts forecasting

1. Introduction
Neural networks are quantitative models linking inputs and outputs adaptively in a learning process analogous to that used by the human brain. The networks consist of elementary units, labeled neurons, joined by a set of rules and weights. The units code characteristics, and they appear in layers, the first being the input layer and the last being the output layer. The data under analysis are processed through different layers, with learning taking place through alteration of the weights connecting the units. At the final iteration, the association between the input and output patterns is established. The example pursued to good expository effect in Neural Networks is face recognition patterns.

Research on neural networks has been going on for some time—for example, the perceptron (the first kind of artificial neural network) was built in the 1950s. Interest declined from the 1960s until the 1980s, when it was renewed. Probably, according to the scientific authors, this renewal of interest resulted from the spreading appreciation of error back-propagation, which could correct weights in the hidden layers. Currently, work in the area is vigorous, led by cognitive psychologists, statisticians, engineers, and mathematicians. In the very last years neural networks have also been applied in the field of spare parts forecasting.

2. What are neural networks?
Neural networks are adaptive statistical models based on an analogy with the structure of the brain. They are adaptive in that they can learn to estimate the parameters of some population using a small number of exemplars (one or a few) at a time. They do not differ essentially from standard statistical models. For example, one can find neural network architectures akin to discriminating analysis, principal component analysis, logistic regression, and other techniques. In fact, the same mathematical tools can be used to analyze standard statistical models and neural networks. Neural networks are used as statistical tools in a variety of fields, including psychology, statistics, engineering, econometrics, and even physics. They are used also as models of cognitive processes by neural- and cognitive scientists.

Basically, neural networks are built from simple units, sometimes called neurons by analogy. These units are interlinked by a set of weighted connections. Learning is
usually accomplished by modification of the connection weights. Each unit codes or corresponds to a feature or a characteristic of a pattern that we want to analyze or that we want to use as a predictor. The units are organized in layers.

The first layer is called the input layer, the last one the output layer. The intermediate layers (if any) are called the hidden layers. The information to be analyzed is fed to the neurons of the first layer and then propagated to the neurons of the second layer for further processing. The result of this processing is then propagated to the next layer and so on until the last layer. There are two kinds of NNs: feed-forward or with feed-back. In the last case the information of a neuron can also go back to precedent neurons. Each unit receives some information from other units (or from the external world through some devices) and processes this information, which will be converted into the output of the unit.

The goal of the network is to learn, or to discover, some association between input and output patterns. This learning process is achieved through the modification of the connection weights between units. In statistical terms, this is equivalent to interpreting the value of the connections between units as parameters (e.g., like the values of $a$ and $b$ in the regression equation $y = a + bx$) to be estimated. The learning process specifies the "algorithm" used to estimate the parameters.

In brief, Haykin (1999, p.2) defines neural networks as follows:

“A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:
1. Knowledge is acquired by the network from its environment through a learning process.
2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.”

Fig 1.1 – A simple example of neural network

It is apparent that a neural network derives its computing power through, first, its massively parallel distributed structure and, second, its ability to learn and therefore generalize. Generalization refers to the neural network producing reasonable outputs for inputs not encountered during training (learning). These two information-processing capabilities make it possible for neural networks to solve complex (large-scale) problems that are currently intractable. In practice, however, neural networks cannot provide the solution by working individually. Rather, they need to be integrated into a consistent system engineering approach. Specifically, a complex problem of interest is decomposed into a number of relatively simple tasks, and neural networks are assigned a subset of the tasks that match their inherent capabilities. It is important to recognize, however, that we have a long way to go (if ever) before we can build a computer architecture that mimics a human brain.

The use of neural networks offers the following useful properties and capabilities (Haykin, 1999, p.2-4).

1. Nonlinearity. An artificial neuron can be linear or nonlinear. A neural network, made up of an interconnection of nonlinear neurons, is itself nonlinear. Moreover, the nonlinearity is of a special kind in the sense that it is distributed throughout the network. Nonlinearity is a highly important property, particularly if the underlying physical mechanism responsible for generation of the input signal is inherently nonlinear.

2. Input-Output Mapping. A popular paradigm of learning called learning with a teacher or supervised learning involves modification of the synaptic weights of a neural network by applying a set of labeled training samples or task examples. Each example consists of a unique input signal and a corresponding desired response. The network is presented with an example picked at random from the set, and the synaptic weights (free parameters) of the network are modified to minimize the difference between the desired response and the actual response of the network produced by the input signal in accordance with an appropriate statistical criterion. The training of the network is repeated for many examples in the set until the network reaches a steady state where there are no further significant changes in the synaptic weights. The previously applied training examples may be reapplied during the training session but in a different order. Thus the network learns from the examples by constructing an input-output mapping for the problem at hand. Such an approach brings to mind the study of nonparametric statistical inference, which is a branch of statistics dealing with model-free estimation; the term "nonparametric" is used here to signify the fact that no prior assumptions are made on a statistical model for the input data. Consider, for example, a pattern classification task, where the requirement is to assign an input signal representing a physical object or
event to one of several pre-specified categories (classes). In a nonparametric approach to this problem, the requirement is to "estimate" arbitrary decision boundaries in the input signal space for the pattern-classification task using a set of examples, and to do so without invoking a probabilistic distribution model. A similar point of view is implicit in the supervised learning paradigm, which suggests a close analogy between the input-output mapping performed by a neural network and nonparametric statistical inference.

3. Adaptation capacity. Neural networks have a built-in capability to adapt their synaptic weights to changes in the surrounding environment. In particular, a neural network trained to operate in a specific environment can be easily retrained to deal with minor changes in the operating environmental conditions. Moreover, when it is operating in a non-stationary environment (i.e., one where statistics change with time), a neural network can be designed to change its synaptic weights in real time. The natural architecture of a neural network for pattern classification, signal processing, and control applications, coupled with the adaptive capability of the network, make it a useful tool in adaptive pattern classification, adaptive signal processing, and adaptive control. As a general rule, it may be said that the more adaptive we make a system, all the time ensuring that the system remains stable, the more robust its performance will likely be when the system is required to operate in a non-stationary environment. It should be emphasized, however, that adaptation capacity does not always lead to robustness; indeed, it may do the very opposite. For example, an adaptive system with short time constants may change rapidly and therefore tend to respond to spurious disturbances, causing a drastic degradation in system performance. To realize the full benefits of adaptation capacity, the principal time constants of the system should be long enough for the system to ignore spurious disturbances and yet short enough to respond to meaningful changes in the environment.

4. Evidential Response. In the context of pattern classification, a neural network can be designed to provide information not only about which particular pattern to select, but also about the confidence in the decision made. This latter information may be used to reject ambiguous patterns, should they arise, and thereby improve the classification performance of the network.

5. Contextual Information. Knowledge is represented by the very structure and activation state of a neural network. Every neuron in the network is potentially affected by the global activity of all other neurons in the network. Consequently, contextual information is dealt with naturally by a neural network.

6. Fault Tolerance. A neural network, implemented in hardware form, has the potential to be inherently fault tolerant, or capable of robust computation, in the sense that its performance degrades gracefully under adverse operating conditions. For
example, if a neuron or its connecting links are damaged, recall of a stored pattern is impaired in quality. However, due to the distributed nature of information stored in the network, the damage has to be extensive before the overall response of the network is degraded seriously. Thus, in principle, a neural network exhibits a graceful degradation in performance rather than catastrophic failure. There is some empirical evidence for robust computation, but usually it is uncontrolled. In order to be assured that the neural network is in fact fault tolerant, it may be necessary to take corrective measures in designing the algorithm used to train the network.

7. Uniformity of Analysis and Design. Basically, neural networks enjoy universality as information processors. We say this in the sense that the same notation is used in all domains involving the application of neural networks. This feature manifests itself in different ways:

- Neurons, in one form or another, represent an ingredient common to all neural networks.
- This commonality makes it possible to share theories and learning algorithms in different applications of neural networks.
- Modular networks can be built through a seamless integration of modules.

8. Neurobiological Analogy. The design of a neural network is motivated by analogy with the brain, which is a living proof that fault tolerant parallel processing is not only physically possible but also fast and powerful. Neurobiologists look to (artificial) neural networks as a research tool for the interpretation of neurobiological phenomena. On the other hand, engineers look to neurobiology for new ideas to solve problems more complex than those based on conventional hard-wired design techniques.

4. Models of a neuron

A neuron is an information-processing unit that is fundamental to the operation of a neural network. The block diagram of Fig. 4.1 shows the model of a neuron, which forms the basis for designing (artificial) neural networks. Here we identify three basic elements of the neuronal model:

1. A set of synapses or connecting links, each of which is characterized by a weight or strength of its own. Specifically, a signal \( x_j \) at the input of synapse \( j \) connected to neuron \( k \) is multiplied by the synaptic weight \( w_{jk} \). It is important to make a note of the manner in which the subscripts of the synaptic weight \( w_{jk} \) are written. The first subscript refers to the neuron in question and the second subscript refers to the input end of the synapse to which the weight refers. Unlike a synapse in the brain, the synaptic weight of an artificial
neuron may lie in a range that includes negative as well as positive values.

2. An adder for summing the input signals, weighted by the respective synapses of the neuron; the operations described here constitute a linear combiner.

3. An activation function for limiting the amplitude of the output of a neuron. The activation function is also referred to as a squashing function in that it squashes (limits) the permissible amplitude range of the output signal to some finite value. Typically, the normalized amplitude range of the output of a neuron is written as the closed unit interval [0,1] or alternatively [-1,1].

![Figure 4.1 – Nonlinear model of a neuron](image)

The neuronal model of Fig. 4.1 also includes an externally applied bias, denoted by $b_k$. The bias $b_k$ has the effect of increasing or lowering the net input of the activation function, depending on whether it is positive or negative, respectively.

In mathematical terms, we may describe a neuron $k$ by writing the following pair of equations:

$$u_k = \sum_{j=1}^{m} w_{kj} \cdot x_j \quad (4.1)$$

and

$$y_k = \rho(u_k + b_k) \quad (4.2)$$

where $x_1, x_2, \ldots, x_m$ are the input signals; $w_{k1}, w_{k2}, \ldots, w_{km}$ are the synaptic weights of neuron $k$; $u_k$ is the linear combiner output due to the input signals; $b_k$ is the bias; $\rho(\cdot)$ is the activation function; and $y_k$ is the output signal of the neuron. The use of bias $b_k$ has the effect of applying an affine transformation to the output $u_k$ of the linear combiner in the
model of Fig. 4.1, as shown by:

\[ v_k = u_k + b_k \]  \hspace{1cm} (4.3)

Bias \( b_k \) is not always present and sometimes it is disguised as input signal.

4.1. Types of activation function

The activation function, denoted by \( \rho(v_k) \), defines the output of a neuron in terms of the induced local field \( v \). Here we identify three basic types of activation functions:

1. **Threshold Function.** For this type of activation function, described in Fig. 4.2a, we have:

\[
\rho(v_k) = \begin{cases} 
1 & \text{if } v_k \geq 0 \\
0 & \text{if } v_k < 0 
\end{cases}  \hspace{1cm} (4.4)
\]

In engineering literature, this form of a threshold function is commonly referred to as a Heaviside function. Correspondingly, the output of neuron \( k \) employing such a threshold function is expressed as:

\[
y_k = \begin{cases} 
1 & \text{if } v_k \geq 0 \\
0 & \text{if } v_k < 0 
\end{cases}  \hspace{1cm} (4.5)
\]

Such a neuron is referred to in the literature as the McCulloch-Pitts model. In this model, the output of a neuron takes on the value of 1 if the induced local field of that neuron is nonnegative, and 0 otherwise. This statement describes the all-or-none property of the McCulloch-Pitts model.

2. **Piecewise-Linear Function.** For the piecewise-linear function described in Fig. 4.2b we have:

\[
\rho(v_k) = \begin{cases} 
1 & v_k \geq +\frac{1}{2} \\
\frac{1}{2} & -\frac{1}{2} < v_k < +\frac{1}{2} \\
0 & v_k \leq -\frac{1}{2} 
\end{cases}  \hspace{1cm} (4.6)
\]

where the amplification factor inside the linear region of operation is assumed to be unity. This form of activation function may be viewed as an approximation to a nonlinear amplifier. The following two situations may be viewed as special forms of the piecewise-linear function:

- A linear combiner arises if the linear region of operation is maintained without running into saturation.
• The piecewise-linear function reduces to a threshold function if the amplification factor of the linear region is made infinitely large.

3. **Sigmoid Function.** The sigmoid function, whose graph is s-shaped (Fig 4.2c), is by far the most common form of activation function used in the construction of artificial neural networks. It is defined as a strictly increasing function that exhibits a graceful balance between linear and nonlinear behavior. An example of the sigmoid function is the logistic function, defined by:

\[
\rho(v_k) = \frac{1}{1 + e^{-av}}
\]

where \( a \) is the slope parameter of the sigmoid function. By varying the parameter \( a \), we obtain sigmoid functions of different slopes, as illustrated in Fig. 4.2c. In the limit, as the slope parameter approaches infinity, the sigmoid function becomes simply a threshold function. Whereas a threshold function assumes the value of 0 or 1, a sigmoid function assumes a continuous range of values from 0 to 1. Note also that the sigmoid function is differentiable, whereas the threshold function is not.

The activation functions defined in Eqs. (4.4), (4.6), and (4.7) range from 0 to +1. It is sometimes desirable to have the activation function range from -1 to +1, in which case the activation function assumes an anti-symmetric form with respect to the origin; that is, the activation function is an odd function of the induced local field. Specifically, the threshold function of Eq. (1.8) is now defined as:

\[
\rho(v_k) = \begin{cases} 
1 & v_k > 0 \\
0 & v_k = 0 \\
-1 & v_k < 0 
\end{cases}
\]

which is commonly referred to as the **signum function.** For the corresponding form of a sigmoid function we may use the **hyperbolic tangent function,** defined by:

\[
\rho(v_k) = \tanh(v_k)
\]
5. Network architectures

The manner in which the neurons of a neural network are structured is intimately linked with the learning algorithm used to train the network. The classification of learning algorithms is considered in the next paragraph. This section is focused on network architectures (structures).

In general, three fundamentally different classes of network architectures may be identified:

1. Single-Layer Feed-forward Networks. In a layered neural network the neurons are organized in the form of layers. In the simplest form of a layered network, we have an input layer of source nodes that projects onto an output layer of neurons (computation nodes), but not vice versa. In other words, this network is strictly a feed-forward or acyclic type. It is illustrated in Fig. 5.1 for the case of four nodes in both the input and output layers. Such a network is called a single-layer network, with the designation "single-layer" referring to the output layer of computation nodes (neurons). We do not count the input layer of source nodes because no computation is performed there.
2. Multilayer Feed-forward Networks. The second class of a feed-forward neural network distinguishes itself by the presence of one or more hidden layers, whose computation nodes are correspondingly called hidden neurons or hidden units. The function of hidden neurons is to intervene between the external input and the network output in some useful manner. By adding one or more hidden layers, the network is enabled to extract higher-order statistics. In a rather loose sense the network acquires a global perspective despite its local connectivity due to the extra set of synaptic connections and the extra dimension of neural interactions. The ability of hidden neurons to extract higher-order statistics is particularly valuable when the size of the input layer is large.

The source nodes in the input layer of the network supply respective elements of the activation pattern (input vector), which constitute the input signals applied to the neurons (computation nodes) in the second layer (i.e., the first hidden layer). The output signals of the second layer are used as inputs to the third layer, and so on for the rest of the network. Typically the neurons in each layer of the network have as their inputs the output signals of the preceding layer only. The set of output signals of the neurons in the output (final) layer of the network constitutes the overall response of the network to the activation pattern supplied by the source nodes in the input (first) layer. The architectural graph in Fig. 5.2 illustrates the layout of a multilayer feed-forward neural network for the case of a single hidden layer. For brevity the network in Fig. 5.2 is referred to as a 10-4-2 network because it has 10 source nodes, 4 hidden neurons, and 2 output neurons. The neural network in Fig. 5.2 is said to be fully connected in the sense that every node in each layer of the network is connected to every other node in the adjacent forward layer. If, however, some of the communication links (synaptic connections) are missing from the network, we say that the network is partially connected.
3. Recurrent Networks. A recurrent neural network distinguishes itself from a feed-forward neural network in that it has at least one feedback loop. For example, a recurrent network may consist of a single layer of neurons with each neuron feeding its output signal back to the inputs of all the other neurons. A recurrent neural network may have or not hidden layers and may have or not self-feedback; self-feedback refers to a situation where the output of a neuron is fed back into its own input. In Fig. 5.3 a class of recurrent networks with hidden neurons is illustrated.

The presence of feedback loops, has a profound impact on the learning capability of the network and on its performance. Moreover, the feedback loops involve the use of particular branches composed of unit-delay elements (denoted by $z^{-1}$), which result in a nonlinear dynamical behavior, assuming that the neural network contains nonlinear units.
6. Learning processes

The property that is of primary significance for a neural network is the ability of the network to learn from its environment, and to improve its performance through learning. The improvement in performance takes place over time in accordance with some prescribed measure. A neural network learns about its environment through an interactive process of adjustments applied to its synaptic weights and bias levels. Ideally, the network becomes more knowledgeable about its environment after each iteration of the learning process.

There are too many activities associated with the notion of "learning" to justify defining it in a precise manner. Moreover, the process of learning is a matter of viewpoint, which makes it all the more difficult to agree on a precise definition of the term. For example, learning as viewed by a psychologist is quite different from learning in a classroom sense. Haykin (1999, p.50) defines learning in the context of neural networks as:

“Learning is a process by which the free parameters of a neural network are adapted through a process of stimulation by the environment in which the network is embedded. The type of learning is determined by the manner in which the parameter changes take place.”

This definition of the learning process implies the following sequence of events:
1. The neural network is stimulated by an environment.
2. The neural network undergoes changes in its free parameters as a result of this stimulation.
3. The neural network responds in a new way to the environment because of the changes that have occurred in its internal structure.

A prescribed set of well-defined rules for the solution of a learning problem is called a learning algorithm. As one would expect, there is no unique learning algorithm for the design of neural networks. Rather, there is a "kit of tools" represented by a diverse variety of learning algorithms, each of which offers advantages of its own. Basically, learning algorithms differ from each other in the way in which the adjustment to a synaptic weight of a neuron is formulated.

In general, we can identify three kinds of learning processes. These are supervised learning, unsupervised learning and reinforcement learning. Usually any given type of network architecture can be employed in any of those tasks.

- supervised learning: in case of disposition of a training set (a set with typical examples of inputs and relative outputs) the network can learn the relation that ties them together. Through an appropriate algorithm that utilizes this training set in order to modify weights and others parameters of the network minimizing the error, the network is trained. If train has success, network learns the
unknown relation between inputs and outputs, and therefore is able to forecast whereas the output is unknown. To do this network has to have an appropriate capacity of generalization, with reference at unknown cases.
- unsupervised learning: it is based on training algorithms that modify weights and other parameters of the network using a dataset of only inputs. These algorithms try to group together the inputs and to identify appropriate clusters, using typically topological or probabilistic methods.
- reinforcement learning: in this learning process an appropriate algorithm has the aim to identify a certain modus operandi, starting from the observation of the external setting; every action has an environmental impact and the environment produces a feedback that directs the algorithm to the learning process. The algorithm "sends" an agent with perception ability in the setting, which explores the environment where it undertakes a series of actions; the environment gives, in response, an incentive or a disincentive. Finally algorithms for the reinforcement learning try to find a policy that maximizes incentives and minimizes disincentives.

It is obvious that in the field of spare parts forecasting, first kind of learning processes is the most used. This because the analysis of the forecast of the spare parts demand starts from a dataset of inputs correlated to relative outputs (that is, in general, the demand of the period).

One of the learning algorithms that have gained more success than others in recent scientific algorithms is back-propagation algorithm (naturally, a supervised learning process). The theory of BP algorithm is explained in the following paragraph, while in paragraph 7.1 a numerical application is given.

6.1. The back-propagation algorithm
Back-propagation algorithm applies a supervised learning. In this explanation, a network with one hidden layer is considered and the following symbols are used:

\( X_k \) is k-th input

\( t_k \) is k-th output

\( Z \) is the matrix of the weights between input and hidden layer

\( W \) is the matrix of the weights between hidden and output layer

\( \eta \) is the learning rate, better explained in paragraph 6.1.1.

This algorithm involves two phases. In the first phase, a forward flow of activation is generated from the input layer to the output layer via the hidden layer. Each unit in the hidden layer computes its activation as a weighted sum of its inputs and transforms it into its response using its transfer function. The \( L \times 1 \) vector of the hidden cell responses is obtained as:
The response of the hidden layer is the input of the output layer. The response vector of the output units is given by:

\[ t'_k = f(W^T \ast h_k) \] (6.2)

In the second phase, the error term, defined as the difference between the actual output and the desired output, is computed. The output cell error term vector is:

\[ e_k = (t_k - t'_k) \] (6.3)

The error term is then transformed into an error signal which takes into account the derivative of the cell activations. The error signal vector, denoted \( \delta_{\text{output},k} \) for the output layer is given by:

\[ \delta_{\text{output},k} = f'(W^T \ast h_k)(e_k) = f'(W^T \ast h_k)(t_k - t'_k) \] (6.4)

where \( f' \) represents the derivative of the transfer function \( f \) and \( \cdot \) the element-wise product of the vectors. When \( f \) is the logistic function, Eq. 6.4 reduces to:

\[ \delta_{\text{output},k} = t'_k(1 - t'_k)(t_k - t'_k) \] (6.5)

where 1 is a unit vector of the same dimensions as \( t_k \) (i.e., a \( J \times 1 \) vector whose elements are 1's).

The error signal is then back-propagated through the network, layer by layer. After the connection weights have been used to back-propagate the error, they are adjusted so as to minimize the mean-square error between the network output and the desired output. The weights in the matrix \( W \) are changed iteratively. For the next iteration the matrix is computed as:

\[ W_{(n+1)} = W_{(n)} + \eta \ast h_k \ast \delta_{\text{output},k}^T \] (6.6)

where \( k \) is randomly chosen.

The adjustment of the weights between the input units and the hidden units (i.e., the weights in matrix \( Z \)) is proportional to both the (estimated) error:

\[ e'_k = W_{(n)} \ast \delta_{\text{output},k} \] (6.7)

of each hidden unit and the extent to which each specific input unit contributed to this error. The (estimated) error signal vector for the hidden units is denoted \( \delta_{\text{hidden},k} \) and is obtained as a weighted combination of the output cell error signals multiplied by the derivative of the hidden cell activations:

\[ \delta_{\text{hidden},k} = f'(Z^T \ast X_k)(e_k) = f'(Z^T \ast X_k)(W_{(n)} \ast \delta_{\text{output},k}) \] (6.8)

With the logistic function, Eq. 6.8 reduces to:

\[ \delta_{\text{hidden},k} = h_k(1 - h_k)(W_{(n)} \ast \delta_{\text{output},k}) \] (6.9)

Finally, learning at iteration \( n+1 \), for the cells of the hidden layer, is implemented as:

\[ Z_{(n+1)} = Z_{(n)} + \eta \ast X_k \ast \delta_{\text{hidden},k}^T \] (6.10)
6.1.1. Rate of learning
The back-propagation algorithm provides an "approximation" to the trajectory in weight space computed by the method of steepest descent. The smaller we make the learning-rate parameter $\eta$, the smaller the changes to the synaptic weights in the network will be from one iteration to the next, and the smoother will be the trajectory in weight space. This improvement, however, is attained at the cost of a slower rate of learning. If, on the other hand, we make the learning-rate parameter $\eta$ too large in order to speed up the rate of learning, the resulting large changes in the synaptic weights assume such a form that the network may become unstable (i.e., oscillatory).

7. Neural networks in spare parts forecasting
The particular features explained in the theory of artificial neural networks have conduced a lot of scientific authors, in the very last years, to apply neural networks in the field of spare parts demand forecasting. The capacity of learning from the environment, the adaptation capacity, the non-linearity are three of the main reasons that have conduced to this application.

Results that have been reached by applying NNs in the field of spare parts are, in general, very satisfactory; however, it's not reasonable to assert that NNs always performs better than other methods. In this work two different case-study are discussed in order to individuate esteems and defects of NNs in this field.

In recent scientific literature, different kinds of neural networks have been applied to spare parts forecasting; in particular, NNs with supervised learning are the most used NNs in this field. The reason is simple: in the field of spare parts forecasting, a data sets of inputs associated to output (the forecast) is, generally, the point start; therefore an algorithm that compares the output of a NN with the actual values of demand and after corrects the weights, is the ideal in this field.

The following paragraph shows a numerical example of applying NNs on spare parts forecasting: the BPN is one of the NNs that have been more discussed and applied than others in the recent literature.

7.1. A numerical example
In order to illustrate the algorithm described in the previous section, a simple numerical example of the application of back-propagation network (BPN) in spare parts demand forecast is given. The BPN used in this example is a three-layer network made of $I = 3$ input units, $L = 2$ hidden units, and $J = 1$ output unit. The input units may be:
- the value of the demand in the last period;
- the intervals (number of periods) between the two last non-zero demand;
- the number of consecutive periods with demand transaction, immediately preceding target period.

While the output unit is the forecast for the target period.

The BPN is trained to associate the stimulus \( i = (1 \ 2 \ 3)^T \) to the response \( t = (1) \) that is the actual number of spare parts required in the target period. The initial connection weights are illustrated in the following figure.

**Figure 7.1 – Network architecture of the example. The input values, output targets and connection weights are indicated on the figure**

The weight matrix \( Z \) connecting the input layer to the hidden layer is an \( I \times L = 3 \times 2 \) matrix. The weight matrix \( W \) connecting the hidden layer to the output layer is an \( L \times J = 2 \times 1 \) matrix. These matrix are equals to:

\[
Z = \begin{bmatrix}
0.5 & 0.3 \\
0.3 & 0.2 \\
0.1 & 0.1 \\
\end{bmatrix}, \quad W = \begin{bmatrix}
0.3 \\
0.4 \\
\end{bmatrix}
\]

The algorithm starts by forwarding the input activation to the output units via the hidden units. First, the input units compute their level of activation (see Fig 7.2), denoted by, as:

\[
b = Z^T \times i = \begin{bmatrix}
0.5 & 0.3 & 0.1 \\
0.3 & 0.2 & 0.1 \\
\end{bmatrix} \times \begin{bmatrix}
1 \\
2 \\
3 \\
\end{bmatrix} = \begin{bmatrix}
1.4 \\
1.0 \\
\end{bmatrix}
\]
Figure 7.2 – First step of the back-propagation algorithm. The stimulus $i$ is presented to the input units which propagate their activation to the hidden units via the first set of connections.

Next, this activation is transformed into a response by using the logistic function (see Fig. 7.3):

$$h = f(b) = f \begin{bmatrix} 1.4 \\ 1.0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 + e^{-1.4} \end{bmatrix} = \begin{bmatrix} 0.8022 \\ 0.7311 \end{bmatrix}$$

Figure 7.3 – The hidden units transform their activation into a response $h$ using the logistic function.

The response is then forwarded to the output units, which compute their activation (see Fig. 7.4):

$$a = W^T \times h = \begin{bmatrix} 0.3 & 0.4 \end{bmatrix} \times \begin{bmatrix} 0.8022 \\ 0.7311 \end{bmatrix} = [0.5331]$$
and transform it into a response using the logistic function (Fig. 7.5):

\[ t' = f(a) = f(0.5331) = \frac{1}{1 + e^{-0.5331}} = 0.6302 \]

![Diagram](image)

**Figure 7.4** – The output units compute their activation \( a \) as the weighted sum of their input.

![Diagram](image)

**Figure 7.5** – The output units transform their activation into a response \( t' \) via the logistic function and propagate it to the supervisor, which computes the error \( e \).

The first step is now accomplished, and learning can begin. The error is computed by the “supervisor” as the difference between the computed response \( t' \) and the expected response \( t \) (see Fig. 7.5):

\[ e = t - t' = 1 - 0.6302 = 0.3698 \]

To compute the error signal \( \delta_{\text{output}} \), the first step is to compute the derivative of the output unit response:

\[ f'(a) = t'(1 - t') = [0.6302] \cdot ([1] - [0.6302]) = [0.2330] \]
the error signal is then computed as the element-wise product of this derivative and the error (see Fig. 7.6):
\[
\delta_{\text{output}} = f'(a) \cdot e = [0.2330, 0.3698] = [0.0862]
\]

The output units will now back-propagate this error signal to the hidden units. First, the amount of the output error attributable to each hidden unit is estimated by multiplying the error signal of the output unit by the weights in W, which connect the output layer to the hidden layer. This propagation of the output error signal is illustrated in Fig. 7.7.

\[
e' = W \times \delta_{\text{output}} = \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix} \times [0.0862] = \begin{bmatrix} 0.0259 \\ 0.0345 \end{bmatrix}
\]

The error signal of the hidden units is then computed similarly to the error signal of the output units, except that the error given by the supervisor is replaced by the estimation of the hidden layer error.
First the derivative if the hidden unit responses is computed:

\[ f'(b) = h \cdot (1 - h) = \begin{bmatrix} 0.8022 \\ 0.7311 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0.8022 \\ 0.7311 \end{bmatrix} = \begin{bmatrix} 0.1587 \\ 0.1966 \end{bmatrix} \]

and than the error signal is given by:

\[ \delta_{\text{hidden}} = f'(b) \cdot e' = \begin{bmatrix} 0.1587 \\ 0.1966 \end{bmatrix} \cdot \begin{bmatrix} 0.0259 \\ 0.0345 \end{bmatrix} = \begin{bmatrix} 0.0041 \\ 0.0068 \end{bmatrix} \]

Fig. 7.8 illustrates this computation of the hidden unit error signal.

Once the error signal is back-propagated to the hidden layer (through the weight matrix \( W \)), the synaptic weights are corrected.

Fig. 7.9 illustrates the steps to correct the weight matrix between the units of the input and hidden layers. The weight matrix \( Z \) can be corrected as \( Z_{(n+1)} \) (to simplify learning rate \( \eta \) is fixed equal to 1):

\[
Z_{(n+1)} = Z + \Delta Z = Z + \eta \cdot i \cdot \delta_{\text{hidden}}^T = \\
= \begin{bmatrix} 0.5 & 0.3 \\ 0.3 & 0.2 \\ 0.1 & 0.1 \end{bmatrix} + 1.0 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 0.0041 & 0.0068 \\ 0.0041 & 0.0068 \\ 0.0041 & 0.0068 \end{bmatrix} = \\
= \begin{bmatrix} 0.5 \cdot 0.0041 + 0.0082 + 0.0123 & 0.3 \cdot 0.0068 + 0.0136 + 0.0204 \\ 0.1 \cdot 0.0041 + 0.0082 + 0.0123 & 0.1 \cdot 0.0068 + 0.0136 + 0.0204 \end{bmatrix} = \\
= \begin{bmatrix} 0.5041 & 0.3068 \\ 0.3082 & 0.2136 \\ 0.1123 & 0.1204 \end{bmatrix} 
\]
The hidden units update the weight matrix

The output unit also corrects his synaptic weight. The weight matrix $W$ becomes $W_{(n+1)}$:

$$W_{(n+1)} = W + \Delta W = W + \eta \cdot h_x \delta^T_{output} =$$

$$= \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix} + 1 \cdot \begin{bmatrix} 0.8022 \\ 0.7311 \end{bmatrix} \times \begin{bmatrix} 0.0862 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.0691 \\ 0.0630 \end{bmatrix} = \begin{bmatrix} 0.3691 \\ 0.4630 \end{bmatrix}$$

The figure 7.10 illustrates the final result of this iteration of the algorithm. These steps are repeated, using the corrected weight matrices, until the output responses are perfect or until some other criterion is met.

**Figure 7.10 – The final configuration with the corrected weights**

8. Inputs and outputs of a neural network for spare parts forecasting

While the output of a neural network which forecasts spare parts demand is generally only one and is the forecasted value, the inputs might be numerous. The common
inputs are: last historical data, mean of historical data, number of zero-demand in last periods and so on with data that look to the past consumption. But it is really important to underline that also other kinds data might be inputs and this is an enormous advantage of neural networks: considering also data as wear and tear (measured by sensors), situation of the maintenance, number of waste products and so on, with NNs there is, in fact, the possibility to have a more detailed visual of variables (that influence the forecasted value) than the other methods.

9. Two cases of study
In order to contextualize the problem of forecasting spare parts and applying neural networks in this field, two case-study are now studied and discussed. In particular, the first shows the defects of neural networks and the second the esteems of neural network forecasting.

9.1. Case-study 1: forecasting CSP in a semiconductor factory
This is the case-study of a semiconductor factory in Taiwan and is presented by Chen et al. (2009, p.228-230).

The critical spare parts in semiconductor are considerably expensive, the purchasing lead time is long, the demand variation is huge, and indispensably play important roles in factory operation. The prices of critical spare parts (CSP) are range from tens to hundreds of thousand dollars. As the equipments operate, some critical spare parts need to be replaced due to wear and tear. If appropriate amount of critical spare parts are not prepared, machines may not be able to function, thus resulting in a waste of resources. However, estimation of the critical spare parts consumption is a complicated subject. In addition to preparing the required CSP of the machines need according to the work orders, there are also other unpredictable factors, such as human factors or spare parts quality problems. Such a circumstance is more obvious in semiconductor industries. For this consideration, it is important to be able to effectively predict the required number of critical spare parts in advance.

This investigation focuses on forecasting the critical spare parts and evaluating the prediction performance of different forecasting methods. Grey prediction model, back-propagation network (BPN) and moving average method (MA) are used to perform CSP demand prediction, so as to effectively predict the required number of CSP which can be provide as a reference of spare parts control.

This investigation is verified by comparing the predicted demand and actual demand of critical spare parts in a semiconductor factories. This company is one of the leading semiconductor factories in Taiwan.
The BGA socket is one of the critical spare parts in this company, which has the characteristics of expensive, large variation of demand, long purchasing lead time and is necessary to the operation of machine (it has been valuated as critical after an AHP). Such condition makes the managers difficult to prepare the required number of BGA sockets. Therefore, this investigation is targeted at the prediction of BGA sockets requirement monthly, and the prediction is carried out using GM(1,1), BPN and MA. As for data collection, the historical requirements of the BGA sockets and the relevant factors in duration of 28 months from September, 2005 to December, 2007 are collected, the last ten months of BGA sockets requirement are used to compare the prediction accuracy of each forecasting method. The benchmark used in order to compare the different forecasting methods is the "average prediction accuracy", which is simply equal to 1-MAPE.

\[ P.A. = 1 - MAPE = 1 - \frac{\sum_{i=1}^{N} |T_i - A_i|}{N} \] (8.1)

9.1.1. Moving average method prediction result

As consider the length of data, the author used 2 to 18 periods of MA to derive the forecasted value of last ten terms of BGA sockets requirement, and compare the difference with the actual requirement. The average prediction accuracy of the MA is shown as Table 9.1.

<table>
<thead>
<tr>
<th>MOVING AVERAGE</th>
<th>MOVING AVERAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Prediction accuracy (%)</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------</td>
</tr>
<tr>
<td>2</td>
<td>66.29</td>
</tr>
<tr>
<td>3</td>
<td>66.59</td>
</tr>
<tr>
<td>4</td>
<td>62.95</td>
</tr>
<tr>
<td>5</td>
<td>61.23</td>
</tr>
<tr>
<td>6</td>
<td>62.28</td>
</tr>
<tr>
<td>7</td>
<td>60.85</td>
</tr>
<tr>
<td>8</td>
<td>60.5</td>
</tr>
<tr>
<td>9</td>
<td>61.92</td>
</tr>
<tr>
<td>10</td>
<td>62.86</td>
</tr>
</tbody>
</table>

Table 9.1 – The average prediction accuracy of MA (n is the number of periods considered in the MA formula)
According to Table 9.1, the prediction average accuracy of all period of MA is 64.28%, and the prediction accuracy of 3-period of MA is 66.59% which has better predict performance than other periods of MA, the result also indicate that the forecasting of CSP requirement is very difficult, not only because of the large data variation, but also the historical data might not enough to predict future demand accurately.

9.1.2. Grey prediction result
This investigation utilizes 4 to 6 entry of GM(1,1) to predict the consumption of BGA socket. The reason is that, according to the 28 months of data length, the GM(1,1) needs at least four data sets to predict future situation, and the more entries of GM(1,1) may not indicate better prediction performance. The average prediction accuracy of the GM(1,1) is presented as Table 9.2.

<table>
<thead>
<tr>
<th>TERM</th>
<th>Actual value</th>
<th>4-entry GM(1,1) predicted value</th>
<th>5-entry GM(1,1) predicted value</th>
<th>6-entry GM(1,1) predicted value</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>202</td>
<td>149.06</td>
<td>161.96</td>
<td>150.85</td>
</tr>
<tr>
<td>20</td>
<td>194</td>
<td>222.27</td>
<td>226.05</td>
<td>227.62</td>
</tr>
<tr>
<td>21</td>
<td>143</td>
<td>244.87</td>
<td>225.01</td>
<td>233.46</td>
</tr>
<tr>
<td>22</td>
<td>224</td>
<td>129.9</td>
<td>174.9</td>
<td>177.09</td>
</tr>
<tr>
<td>23</td>
<td>184</td>
<td>223.16</td>
<td>194.9</td>
<td>220.83</td>
</tr>
<tr>
<td>24</td>
<td>143</td>
<td>223.65</td>
<td>199.47</td>
<td>187.54</td>
</tr>
<tr>
<td>25</td>
<td>137</td>
<td>116.24</td>
<td>167.96</td>
<td>160.96</td>
</tr>
<tr>
<td>26</td>
<td>139</td>
<td>111.71</td>
<td>106.79</td>
<td>142.62</td>
</tr>
<tr>
<td>27</td>
<td>68</td>
<td>135.67</td>
<td>116.86</td>
<td>106.23</td>
</tr>
<tr>
<td>28</td>
<td>276</td>
<td>65.3</td>
<td>79.5</td>
<td>78.81</td>
</tr>
<tr>
<td>Average accuracy</td>
<td>55.76%</td>
<td>65.23%</td>
<td>67.42%</td>
<td></td>
</tr>
</tbody>
</table>

Table 9.2 – The prediction result of 4, 5, 6-entry of GM (1,1)

The Table 9.2 shows the 6-entry of GM(1,1) with an average accuracy of 67.42% is higher than 4 and 5 entries of GM(1,1). In this case, the most suitable entry of GM(1,1) is 6. It might imply that when managers decide the demand quantities of BGA sockets, they should consider six months of historical data at least.
9.1.3. BPN result

This investigation also applies back – propagation network to predict the value of BGA sockets in last ten terms.

At the training and testing process of BPN, the first 18 data sets are used for training process, and the last 10 data sets are used for testing process. The suitable parameters setting of the BPN is derived by trial and error. The parameters setting and the prediction accuracy of BPN are listed in Table 9.3.

<table>
<thead>
<tr>
<th>Leaning rule</th>
<th>BPN algorithm</th>
<th>Hidden nodes</th>
<th>Learning rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trans. function</td>
<td>Sigmoid function</td>
<td>Learning rate</td>
<td>1.0</td>
</tr>
<tr>
<td>Input nodes</td>
<td>5</td>
<td>N. of training cycle</td>
<td>1000</td>
</tr>
<tr>
<td>Output nodes</td>
<td>1</td>
<td>Average accuracy</td>
<td>66.02%</td>
</tr>
<tr>
<td>N. of hidden layer</td>
<td>1</td>
<td>MAPE</td>
<td>0.3398</td>
</tr>
</tbody>
</table>

*Table 9.3 – Parameter setting and prediction accuracy*

Because of the demand of BGA sockets doesn’t present null values, the following parameter have been used as input nodes:
- the demand at the end of the immediately preceding target period;
- the demand at the end of two periods before target period;
- the mean of demand for four periods immediately preceding target period;
- the maximum demand among four periods immediately preceding target period;
- the minimum demand among four periods immediately preceding target period;

According to the Table 3, the average accuracy of BPN is 66.02%, and the MAPE of BPN is 0.3398.

9.1.4. Conclusions

Table 9.4 shows the results of the investigation.

<table>
<thead>
<tr>
<th>Forecasting methods</th>
<th>Average prediction accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA (3-period)</td>
<td>66.59</td>
</tr>
<tr>
<td>GM (1,1) (6-entry)</td>
<td>67.42</td>
</tr>
<tr>
<td>BPN</td>
<td>66.02</td>
</tr>
</tbody>
</table>

*Table 9.4 – The highest prediction accuracy of each forecasting methods*

According to the Table 9.4, the GM(1,1) (6-entry) have higher average accuracy of 67.42% than BPN and MA, the order from high to low average prediction accuracy of prediction methods is GM(1,1) (6-entry), MA (3-period), BPN. It can be clearly understand when the data sets is few, the data variation is large and the value of some
influential factors is unknown at the prediction timing of current term, the GM(1,1) might have better prediction performance than BPN and MA.

In this investigation all the defects the of NNs clearly appears. In fact, in spite of they are considered as the best performing forecasting methods from the majority of the scientific authors, they don’t well perform when data sets is few: a large training set is needed in order to take advantage of their peculiarities. In other cases, also traditional methods (as Moving Average) perform better.

9.2. Case study 2: lumpy demand forecasting for spare parts in a petrochemical factory

This second case-study deals with forecast future demand of spare parts of Arak Petrochemical Company in Iran and is presented by Amin-Naseri and Tabar (2008, p.1379-1381).

In order to establish the superiority of neural networks in forecasting lumpy demand, In this study real data sets of 30 types of spare parts demand in Arak Petrochemical Company in Iran have been used. The data were handled for 67 monthly periods from 2001 to 2006. The data series were divided into two sets; namely training and test sets. From 67 monthly observations, 55 observations have been used for training the networks, and five methods tested using the last 12 observations.

The methods compared were:
- three kinds of neural networks:
  - Multilayer Perceptron Network (MLP)
  - Generalized Regression Neural Network (GRNN)
  - Recurrent Neural Network (RNN)
- Croston’s method
- SBA method

The benchmarks used to compare the results are A-MAPE and PB (only between SBA, MLP and RNN).

9.2.1 Multilayer Perceptron Network

The first neural network adopted in this study is the most widely used method, multilayered perceptron (MLP) trained by back-propagation (BP) algorithm. Amin-Naseri and Tabari uses a three layers MLP : the input layer with two input variables, hidden unit layer, and output layer. The MLP has three nodes in the hidden layer. One output unit is used in the output layer.

The input nodes represent two variables:
1) the demand at the end of the immediately preceding period;
2) the number of periods separating the last two nonzero demand transactions as of the end of the immediately preceding period.
The output node represents the predicted value of the demand transaction for the current period. The learning rate value used is 0.1

9.2.2. Generalized Regression Neural Network
Generalized Regression Neural Network (GRNN), the second network proposed in this study, does not require an iterative training procedure as in back propagation method. The GRNN consists of four layers: the input layer, pattern layer, summation layer and output layer.
The effective input variables for the GRNN are as follows:
1. The demand at the end of the immediately preceding target period.
2. The number of consecutive period with no demand transaction immediately preceding target period.
3. The number of periods separating the last two nonzero demand transactions as of the end of the immediately preceding target period.
4. The mean of demand for four period immediately preceding target periods.

9.2.3. Recurrent Neural Network
The third neural network is a recurrent neural network RNN. The network consists of four layers: an input layer, a hidden layer, a context layer and an output layer.
In this network, following variables have been defined for input nodes in input layer:
1. The demand at the end of the immediately preceding target period.
2. The number of consecutive periods with demand transaction, immediately preceding target period.
3. The number of consecutive period with no demand transaction, immediately preceding target period.
4. The number of periods separating the last two nonzero demand transactions as of the end of the immediately preceding target period.
5. The number of period(s) between target period and first nonzero demand immediately preceding target period.
6. The number of period(s) between target period and first zero demand immediately preceding target period.
7. The mean of demand for six periods immediately preceding target period.
8. The maximum demand among six periods immediately preceding target period.
9.2.4. Croston's method and SBA method

In this study, a wide range of smoothing constants from 0.05 to 0.45 with increments of 0.05 have been used for Croston's method and Syntetos-Boylan approximation (SBA).

9.2.5. Results and conclusions

Tables 8.5 and 8.6 show the results of the investigation. Table 8.5 reports the average of adjusted MAPE of the 30 spare parts forecasts for the methods under consideration and Table 8.6 the model performance based on percentage best statistic (PB).

According to these tables, neural networks perform better than Croston’s methods and SBA. In particular, the order of the best performing methods is: RNN, GRNN, MLP, SBA, and CR.

In reverse to the first case-study, in this case, the esteem of NNs clearly appear: when the dataset for training process is large enough and also when demand is particularly lumpy, the superiority of NNs over classical methods is evident.

<table>
<thead>
<tr>
<th>CR</th>
<th>SBA</th>
<th>MPL</th>
<th>GRNN</th>
<th>RNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>58.64%</td>
<td>49.95%</td>
<td>48.31%</td>
<td>42.08%</td>
<td>30.85%</td>
</tr>
</tbody>
</table>

Table 9.5 – Average A-MAPE for forecasting methods

<table>
<thead>
<tr>
<th>Spare part</th>
<th>SBA</th>
<th>MLP</th>
<th>RNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.67</td>
<td>16.67</td>
<td>66.67</td>
</tr>
<tr>
<td>2</td>
<td>25.00</td>
<td>41.67</td>
<td>33.33</td>
</tr>
<tr>
<td>3</td>
<td>66.67</td>
<td>16.67</td>
<td>16.67</td>
</tr>
<tr>
<td>4</td>
<td>8.33</td>
<td>33.33</td>
<td>58.33</td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td>25.00</td>
<td>75.00</td>
</tr>
<tr>
<td>6</td>
<td>8.33</td>
<td>33.33</td>
<td>58.33</td>
</tr>
<tr>
<td>7</td>
<td>25.00</td>
<td>33.33</td>
<td>41.67</td>
</tr>
<tr>
<td>8</td>
<td>66.67</td>
<td>16.67</td>
<td>16.67</td>
</tr>
<tr>
<td>9</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td>10</td>
<td>8.33</td>
<td>16.67</td>
<td>75.00</td>
</tr>
<tr>
<td>11</td>
<td>16.67</td>
<td>25.00</td>
<td>58.33</td>
</tr>
<tr>
<td>12</td>
<td>8.33</td>
<td>8.33</td>
<td>83.33</td>
</tr>
<tr>
<td>13</td>
<td>41.67</td>
<td>25.00</td>
<td>33.33</td>
</tr>
<tr>
<td>14</td>
<td>25.00</td>
<td>33.33</td>
<td>41.67</td>
</tr>
<tr>
<td>15</td>
<td>25.00</td>
<td>25.00</td>
<td>50.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>16</td>
<td>41.67</td>
<td>8.33</td>
<td>50.00</td>
</tr>
<tr>
<td>17</td>
<td>41.67</td>
<td>8.33</td>
<td>50.00</td>
</tr>
<tr>
<td>18</td>
<td>33.33</td>
<td>8.33</td>
<td>58.33</td>
</tr>
<tr>
<td>19</td>
<td>8.33</td>
<td>58.33</td>
<td>33.33</td>
</tr>
<tr>
<td>20</td>
<td>16.67</td>
<td>16.67</td>
<td>66.67</td>
</tr>
<tr>
<td>21</td>
<td>25.00</td>
<td>33.33</td>
<td>41.67</td>
</tr>
<tr>
<td>22</td>
<td>16.67</td>
<td>41.67</td>
<td>41.67</td>
</tr>
<tr>
<td>23</td>
<td>16.67</td>
<td>41.67</td>
<td>41.67</td>
</tr>
<tr>
<td>24</td>
<td>0.00</td>
<td>8.33</td>
<td>83.33</td>
</tr>
<tr>
<td>25</td>
<td>0.00</td>
<td>8.33</td>
<td>91.67</td>
</tr>
<tr>
<td>26</td>
<td>0.00</td>
<td>25.00</td>
<td>75.00</td>
</tr>
<tr>
<td>27</td>
<td>16.67</td>
<td>16.67</td>
<td>66.67</td>
</tr>
<tr>
<td>28</td>
<td>8.33</td>
<td>16.67</td>
<td>50.00</td>
</tr>
<tr>
<td>29</td>
<td>25.00</td>
<td>33.33</td>
<td>41.67</td>
</tr>
<tr>
<td>30</td>
<td>0.00</td>
<td>25.00</td>
<td>75.00</td>
</tr>
<tr>
<td>Average</td>
<td>19.94</td>
<td>23.60</td>
<td>56.46</td>
</tr>
</tbody>
</table>

Table 9.6 – PB statistics of three forecasting methods
CHAPTER 4
Application of forecasting methods on real dataset

1. Introduction
In this section single exponential smoothing, Croston’s method, Syntetos-Boylan approximation, moving average, weighted moving average, Holt-Winters (no seasonal), ARIMA and a neural network are used to forecast spare parts demand for a business in the iron and steel sector.
Starting from a list of 20 different spare parts with their consumptions in one year, 3 kinds of spare parts were chosen for the analysis. The year was divided into 53 periods (the weeks): the first 38 weeks were used as training set, the last 15 as testing set.
The following table show the demand of these spare parts.

<table>
<thead>
<tr>
<th>WEEK</th>
<th>C010040069</th>
<th>C010270239</th>
<th>C010160001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>114</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>664</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>23</td>
<td>99</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>56</td>
<td>40</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>18</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>32</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>54</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>18</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>15</td>
<td>40</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>20</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>24</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>23</td>
<td>38</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>
C010040069 refers to a black washer, C010270239 refers to a disc grindstone mm 1,5 x 115, while C010160001 refers to a yellow marker. They are more consumption material than spare parts, but they completely fall within the category of spare parts, because their demand is sporadic and tied to the casual phenomenon of breakdown during the functioning of the system where they are used.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>0</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>25</td>
<td>22</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>4</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>28</td>
<td>20</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>320</td>
<td>32</td>
<td>100</td>
</tr>
<tr>
<td>32</td>
<td>81</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>28</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>6</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>8</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>36</td>
<td>0</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>37</td>
<td>120</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>38</td>
<td>8</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>0</td>
<td>53</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>14</td>
<td>16</td>
<td>100</td>
</tr>
<tr>
<td>41</td>
<td>44</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
<td>13</td>
<td>100</td>
</tr>
<tr>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>9</td>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>28</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>47</td>
<td>22</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>48</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>49</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>51</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>52</td>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>53</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ADI</td>
<td>1,529</td>
<td>1,19</td>
<td>3,125</td>
</tr>
<tr>
<td>CV</td>
<td>2,816</td>
<td>1,09</td>
<td>1,676</td>
</tr>
</tbody>
</table>

Table 1.1 – Consumption, ADI and CV of every spare part
These spare parts were chosen among others for the diversity of their ADI and CV. It is evident that CV is very high for all the spare parts under observation, in particular for first and third; for this reason the forecasts will not be much precise.

For exponential smoothing, Croston’s method, Syntetos-Boylan approximation, weighted moving average, Holt-Winters (no seasonal) and ARIMA, testing set was used to individuate the coefficients that minimize the forecast error. For neural network, training set was used to find the right weights with a back-propagation algorithm. In particular, the neural network used has 5 input nodes (the mean, the maximum value, the minimum value of last 5 periods, the last value of consumption and the number of periods with zero-demand in last 5 periods), 3 hidden nodes and 1 output node (the forecast for the next period). The activation function used for hidden and output nodes is the sigmoid function with a=1 and learning rate used is equal to 1. In order to avoid the saturation of the neural network and since the output is between 0 and 1, the inputs 1, 2, 3 and 4 and the output were not the real values but the value divided by the maximum consumption of the training set (if in the testing there was another maximum value it was updated).

To evaluate the performances of each method, the benchmark used is the A-MAPE:

$$\text{A-MAPE} = \frac{\sum_{t=1}^{N} |A_t - F_t|}{N \sum_{t=1}^{N} A_t}$$

In addition to this, other two extenuating circumstances were taken under consideration: ME (Mean Error) and MSE (Mean Squared Error).

The first can be defined:

$$\text{ME} = \frac{1}{N} \sum_{t=1}^{N} (A_t - F_t)$$

The ME is not an accuracy measure as it does not provide information of the forecast errors. A perfect score, ME=0, does not exclude very large and compensating errors of opposite signs. It is also important to remember that a non-zero ME does not necessarily imply a “flat bias”, i.e. a mean error independent of forecast value. ME needs, in particular, to correct the forecasts (by adding it to the forecast value) when $(A_t - F_t)$ is nearly constant in every forecasting period.

The second is the RMSE raised to the second power, it is useful to see how much is great the forecast error when it the forecast is not right.
2. Elaboration of forecasts

For SES, CR, SBA, MA, WMA and NN the forecasts were elaborated by creating appropriate tables in Microsoft Excel; for Holt Winter, SES and ARIMA the software E-views 5.0 was used in order to find the coefficients that minimized the errors of training set, while for SES, CR, SBA different values of the coefficient were tested in order to find the one that minimized the A-MAPE in the training set. For MA and WMA periods from 2 to 5 were tested and for every period of WMA different random weights were tested and those with the minimum A-MAPE in the training set were maintained. For HW and ARIMA E-views 5.0 was also used to elaborate the forecasts; in case of negative value, the forecast was assumed equal to zero. For ARIMA different values of p, d, q were tested and then those with best performances (A-MAPE) were chosen.

The following table shows, for each method, the forecasts elaborated. For SES, CR and SBA the parenthetical number refers to the coefficient used, while for MA and WMA only the period with best results is exposed. All the decimal forecasts were rounded to the closest whole number. First table refers to spare part C010040069, second table to C010270239, and third table to C010160001.

<table>
<thead>
<tr>
<th>WEEK</th>
<th>SES (0.01)</th>
<th>CR (0.003)</th>
<th>SBA (0.1)</th>
<th>MA 4</th>
<th>WMA 5</th>
<th>HW</th>
<th>(4,1,0)</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>34</td>
<td>7</td>
<td>34</td>
<td>14</td>
<td>38</td>
</tr>
<tr>
<td>40</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>18</td>
<td>11</td>
<td>33</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>41</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>8</td>
<td>12</td>
<td>33</td>
<td>23</td>
<td>36</td>
</tr>
<tr>
<td>42</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>17</td>
<td>47</td>
<td>33</td>
<td>34</td>
<td>37</td>
</tr>
<tr>
<td>43</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>15</td>
<td>2</td>
<td>33</td>
<td>28</td>
<td>37</td>
</tr>
<tr>
<td>44</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>6</td>
<td>4</td>
<td>32</td>
<td>26</td>
<td>36</td>
</tr>
<tr>
<td>45</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>15</td>
<td>9</td>
<td>31</td>
<td>28</td>
<td>37</td>
</tr>
<tr>
<td>46</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>9</td>
<td>27</td>
<td>32</td>
<td>27</td>
<td>37</td>
</tr>
<tr>
<td>47</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>13</td>
<td>1</td>
<td>31</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>48</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>15</td>
<td>21</td>
<td>31</td>
<td>28</td>
<td>37</td>
</tr>
<tr>
<td>49</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>13</td>
<td>1</td>
<td>30</td>
<td>28</td>
<td>38</td>
</tr>
<tr>
<td>50</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>6</td>
<td>4</td>
<td>29</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>51</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>6</td>
<td>0</td>
<td>29</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>52</td>
<td>13</td>
<td>9</td>
<td>17</td>
<td>30</td>
<td>2</td>
<td>28</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>53</td>
<td>14</td>
<td>9</td>
<td>16</td>
<td>8</td>
<td>25</td>
<td>28</td>
<td>28</td>
<td>35</td>
</tr>
</tbody>
</table>

*Table 2.1 – Forecasting values of first spare part*
### Table 2.2 – Forecasting values of second spare part

<table>
<thead>
<tr>
<th>WEEK</th>
<th>SES (0,055)</th>
<th>CR (0,03)</th>
<th>SBA (0,51)</th>
<th>MA 2</th>
<th>WMA 5</th>
<th>HW</th>
<th>(0,1,2)</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>11</td>
<td>13</td>
<td>15</td>
<td>22</td>
<td>16</td>
<td>17</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>40</td>
<td>14</td>
<td>14</td>
<td>27</td>
<td>17</td>
<td>26</td>
<td>23</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>41</td>
<td>14</td>
<td>14</td>
<td>20</td>
<td>25</td>
<td>25</td>
<td>23</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>42</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>0</td>
<td>33</td>
<td>21</td>
<td>10</td>
<td>28</td>
</tr>
<tr>
<td>43</td>
<td>14</td>
<td>14</td>
<td>12</td>
<td>12</td>
<td>21</td>
<td>20</td>
<td>10</td>
<td>28</td>
</tr>
<tr>
<td>44</td>
<td>13</td>
<td>14</td>
<td>12</td>
<td>28</td>
<td>22</td>
<td>17</td>
<td>10</td>
<td>23</td>
</tr>
<tr>
<td>45</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>16</td>
<td>15</td>
<td>20</td>
<td>10</td>
<td>23</td>
</tr>
<tr>
<td>46</td>
<td>15</td>
<td>15</td>
<td>18</td>
<td>34</td>
<td>15</td>
<td>23</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>47</td>
<td>14</td>
<td>14</td>
<td>10</td>
<td>15</td>
<td>17</td>
<td>20</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>48</td>
<td>14</td>
<td>14</td>
<td>6</td>
<td>6</td>
<td>15</td>
<td>17</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>49</td>
<td>13</td>
<td>14</td>
<td>6</td>
<td>4</td>
<td>18</td>
<td>16</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>50</td>
<td>13</td>
<td>14</td>
<td>7</td>
<td>9</td>
<td>13</td>
<td>16</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>51</td>
<td>12</td>
<td>14</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>13</td>
<td>10</td>
<td>26</td>
</tr>
<tr>
<td>52</td>
<td>12</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>13</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>53</td>
<td>12</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>11</td>
<td>10</td>
<td>21</td>
</tr>
</tbody>
</table>

### Table 2.3 – Forecasting values of third spare part

<table>
<thead>
<tr>
<th>WEEK</th>
<th>SES (0,001)</th>
<th>CR (0,001)</th>
<th>SBA (0,001)</th>
<th>MA 5</th>
<th>WMA 4</th>
<th>HW</th>
<th>(1,1,0)</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>40</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>80</td>
<td>83</td>
<td>12</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>40</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>42</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>80</td>
<td>17</td>
<td>17</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>43</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>40</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>44</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>40</td>
<td>100</td>
<td>23</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>40</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>46</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>80</td>
<td>83</td>
<td>16</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>47</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>40</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>48</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>17</td>
<td>22</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>49</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>80</td>
<td>83</td>
<td>16</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>51</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>40</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>52</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>17</td>
<td>23</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>53</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>27</td>
</tr>
</tbody>
</table>
Finally, the following tables show the A-MAPE, ME an MSE for each combination of forecasting method-spare part.

<table>
<thead>
<tr>
<th>A-MAPE</th>
<th>SES</th>
<th>CR</th>
<th>SBA</th>
<th>MA</th>
<th>WMA</th>
<th>HW</th>
<th>ARIMA</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>C010040069</td>
<td>1,15</td>
<td>1,06</td>
<td>1,25</td>
<td>1,31</td>
<td>1,39</td>
<td>2,02</td>
<td>2,48</td>
<td>2,45</td>
</tr>
<tr>
<td>C010270239</td>
<td>0,94</td>
<td>0,96</td>
<td>0,84</td>
<td>1,04</td>
<td>1,08</td>
<td>1,13</td>
<td>0,77</td>
<td>1,51</td>
</tr>
<tr>
<td>C010160001</td>
<td>1,03</td>
<td>1,09</td>
<td>1,09</td>
<td>1,65</td>
<td>0,67</td>
<td>1,45</td>
<td>1</td>
<td>1,48</td>
</tr>
</tbody>
</table>

*Table 3.1 – A-MAPE for each combination method-spare part*

<table>
<thead>
<tr>
<th>ME</th>
<th>SES</th>
<th>CR</th>
<th>SBA</th>
<th>MA</th>
<th>WMA</th>
<th>HW</th>
<th>ARIMA</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>C010040069</td>
<td>-2,47</td>
<td>2,47</td>
<td>-4,53</td>
<td>-4,33</td>
<td>-0,67</td>
<td>-31,47</td>
<td>-16,07</td>
<td>-25,53</td>
</tr>
<tr>
<td>C010270239</td>
<td>-3,33</td>
<td>-3,8</td>
<td>-1,4</td>
<td>-2,75</td>
<td>-4</td>
<td>38,53</td>
<td>2,87</td>
<td>-12,87</td>
</tr>
<tr>
<td>C010160001</td>
<td>25,55</td>
<td>22</td>
<td>22</td>
<td>-1,33</td>
<td>0</td>
<td>-20,33</td>
<td>26,67</td>
<td>-0,6</td>
</tr>
</tbody>
</table>

*Table 3.2 – ME for each combination method-spare part*

<table>
<thead>
<tr>
<th>MSE</th>
<th>SES</th>
<th>CR</th>
<th>SBA</th>
<th>MA</th>
<th>WMA</th>
<th>HW</th>
<th>ARIMA</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>C010040069</td>
<td>191,67</td>
<td>165,13</td>
<td>209,33</td>
<td>266,6</td>
<td>411,6</td>
<td>974,53</td>
<td>474,07</td>
<td>855</td>
</tr>
<tr>
<td>C010270239</td>
<td>129,07</td>
<td>133,13</td>
<td>118,33</td>
<td>275,8</td>
<td>245,4</td>
<td>1426</td>
<td>253,93</td>
<td>406,42</td>
</tr>
<tr>
<td>C010160001</td>
<td>2614,87</td>
<td>2423,33</td>
<td>2423,33</td>
<td>2480</td>
<td>1222,3</td>
<td>419,93</td>
<td>2666,67</td>
<td>1970,6</td>
</tr>
</tbody>
</table>

*Table 3.3 – MSE for each combination method-spare part*

### 3. Conclusions

The results of the analysis (A-MAPE) show that when the CV is very high, it is really difficult to find a forecasting method which performs in a good way. In this study, only for the second spare part, whose CV is smaller than others, the result can be considered appreciable; only WMA lead to good results with third spare parts, while for first spare parts, which has the greatest CV, no methods has an A-MAPE smaller than 1. One solution to have better results might be to group together using monthly periods and to have at disposal a dataset of more than one year; another solution might be to use other forecasting methods (for example, other types of Neural networks) which better performs in these cases.

In general, for SES, CR and SBA, when the CV is so high, the coefficients that minimize the errors in the training set are small; the consequence is that the forecasts
are almost constant. Another conclusion appears from table of MSE: the more ADI is high the more the forecast error (when it occurs) is great.

ARIMA method outperforms the other methods for spare part C010270239, for spare parts C010040069 Croston’s method gives the best results, while WMA gives very good forecasts for spare part C010160001. HW method and, in particular, the neural network give bad results by comparison with other methods. For the first, the reason may be that HW is useful when there is a seasonal trend and, in these cases, no seasonal trends can be found. For the second there may be different reasons:

- the use of a neural network too simple for the complexity of the dataset;
- the need of a more extensive training set if the CV is so high;
- the lack of other inputs, such as situation of maintenance of the machine, which let NN outperform other methods.

The following graphs show the forecasts compared to the real dataset.
CONCLUSIONS AND FUTURE RESEARCH

Accurate demand forecasting of spare parts is of importance in inventory management. Owing to demand characteristics of spare parts, demand forecasting in this area is very difficult.

In this work, after a presentation and a contextualization of the problem, I discussed the demand forecasting methods that have been more studied than others in the area of spare parts forecasting. For each methods I showed their innovative features, their limits, I gave a brief explanation and a round-up of the most recent or studied scientific articles. In particular I went into more depth in the study of neural networks and their applicability in spare parts forecasting, also presenting two cases of study. Finally I presented the application of some methods on real dataset.

The work done in this field (in particular in case of lumpy demand) is not so much; subject and research on spare parts management mostly focused on the consideration of safe inventory level and the annexed costs. Chen et al. (2009, p.225) say that if the actual number of spare parts can be correctly predicted, there will be no problem of controlling inventory level and purchasing quantities.

Hence, research in this field is very important and has deep margins of improvement. The directions of future research on spare parts forecasting and, in particular, on the application of neural networks in this field, can be explained in four points (Gutierrez et al., 2008, p.418-419):

1 - To look more closely into factors that lead to a diminution in performance of neural networks models relative to traditional time-series forecasting techniques. The objective will be to identify conditions under which either NN or traditional models would be expected to perform better in forecasting spare parts demand, in particular in case of lumpy demand.

2 - Another interesting research issue is the possibility of combining traditional models with NN models to build hybrid models; combined forecasts are useful when two or more different methods close to actual in different directions. Rules to combine NN models, or other models, with traditional methods can be generated in particular for a typical lumpy demand in future studies.

3 - Improved forecasting accuracy does not necessarily translate into better stock control performance. Syntetos and Boylan (2005, p.498-502), for instance, have empirically assessed stock control performance associated with the use of four traditional time-series forecasting methods (exponential smoothing, simple moving average of length 13, Croston’s method, and SBA). Using simulation, they established the overall superiority of SBA in stock control performance with respect to three possible managerial considerations (a specified customer service level and two cost
policies). For this, another point to develop in the future might be to likewise evaluate and compare stock control performance associated with NN model forecasts in relation to the traditional time-series methods.

4 – Finally, there is the possibility of the introduction of new spare parts demand forecasting methods which perform better.
BIBLIOGRAPHY

Articles:


Books: